MODELLING SPATIAL TRENDS AND LOCAL COMPETITION EFFECTS USING SEMIPARAMETRIC ADDITIVE MODELS

By

Maria L. Durbán Reguera

SUBMITTED FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
AT HERIOT-WATT UNIVERSITY
ON COMPLETION OF RESEARCH IN THE
DEPARTMENT OF ACTUARIAL MATHEMATICS & STATISTICS
NOVEMBER 1998.

I hereby declare that the work presented in this thesis was carried out by myself, except where due acknowledgement is made, and has not been submitted for any other degree.

Maria L. Durbán Reguera (Candidate)

I. D. Currie (Supervisor)

C. A. Hackett (Supervisor)

24 November 1998

Date

Contents

1 Spatial models for the analysis of agricultural trials: an overview **1**

1.1 Design of experiments 1

1.2 Spatial Models 3

1.2.1 Papadakis analysis of covariance 3

1.2.2 Trend as a random process 4

1.3 Non-parametric regression 7

1.3.1 Kernel Methods 8

1.3.2 Smoothing splines 9

1.3.3 Further topics in smoothing 10

1.3.4 Additive models 12

1.3.5 Semiparametric models 14

1.4 Competition Models 16

2 Semiparametric additive models **18**

2.1 Introduction 18

2.2 Additive models 19

2.2.1 Solution to the normal equations 20

2.2.2 Symmetry of the hat matrix 26

2.2.3 Estimation of additive models when all smoothers but one are centred 26

2.2.4 Existence and uniqueness of the solutions 28

2.3 Semiparametric additive models as a special case of additive models . 30

2.3.1 Extension of Speckman's method 32

2.3.2 Estimation of residual variance in SAMs 33
2.4 Standard errors in SAMs 34
2.4.1 The calculation of the standard errors 35
2.4.2 Two approximations for the calculation of the standard errors 35
2.4.3 Some examples 36

3 Exact adjustment of the profile likelihood for a class of normal regression models 44
3.1 Introduction 44
3.2 Notation and definitions 45
3.3 Calculation of the adjusted score 48
3.4 Approximations to the adjusted profile likelihood 55
3.4.1 Cox and Reid approximation 56
3.4.2 McCullagh and Tibshirani approximation 57
3.5 Examples 59
3.5.1 Models with known X 59
3.5.2 Models with independent and identically distributed errors . . 60
3.5.3 A general transformation model 63

4 Smoothing in the presence of a competition effect 70
4.1 Introduction 70
4.2 Competition between plots 71
4.2.1 Models with one smooth term 71
4.2.2 Extension to two or more smooth terms 76
4.3 Competition between neighbouring treatments 78
4.4 Smoothing splines as mixed models and competition 81
4.4.1 Notation 82
4.4.2 Estimation of the random and fixed effects 85
4.4.3 Estimation of the variance components and competition parameter 87
4.5 Smoothing parameter selection 90
4.5.1 Altman's indirect method 91

4.5.2 Wang's modified GCV 91
4.6 Standard errors of treatments in the presence of competition 92
4.7 Examples 93

5 Analysis of cereal cultivar field trials **96**
5.1 Introduction 96
5.2 Strategies for selecting the spatial component of a semiparametric additive model 98
5.2.1 Layout of trial BB 98
5.2.2 Semiparametric modelling 99
5.2.3 Graphical methods 100
5.2.4 Analytical Methods 102
5.2.5 Analysis of trial BB 105
5.2.6 Comparison with other spatial analyses 110
5.3 Competition, trend or correlation?: Analysis of trial BJ 113
5.3.1 Layout of trial BJ 114
5.3.2 Analysis of variance 116
5.3.3 Semiparametric modelling 117

6 Conclusions and further work **127**
6.1 Conclusions 127
6.1.1 Additive and semiparametric models 127
6.1.2 Profile likelihood 128
6.1.3 Semiparametric models and competition 129
6.2 Further work 130
6.2.1 Semiparametric split-plot analysis 130
6.2.2 Profile likelihood 131
6.2.3 Smoothing and mixed models 132

List of Figures

2.1 *Plot of percentage error in the standard error of difference in pairwise comparisons versus degrees of freedom. The mean and maximum percentage for approximation* (2.55) ($u_1(\delta^*)$, $v_1(\delta^*)$) *are represented by* —— *and* ····· , *while the mean and maximum percentage for approximation* (2.56) ($u_2(\delta^*)$, $v_2(\delta^*)$) *are represented by* - - - - - *and* — — — —. *The preferred degrees of freedom is indicated by* △. 42

2.2 *Plot of percentage error in the standard error of difference in pairwise comparisons versus degrees of freedom for the smooth part of the model. The mean and maximum percentage for approximation* (2.56) ($u(d^*)$, $v(d^*)$) *are represented by* - - - - - *and* — — — —. *The preferred degrees of freedom is indicated by* △. 43

3.1 *Plot of the profile log-likelihood —- and adjusted profile log-likelihood - - - for the Draper and Guttman example.* 63

3.2 *Plot of* $tr(PWD^{-1})$ *—- and the quadratic function* $q(\rho)$ *- - - with the same first two derivatives at* $\rho = \hat{\rho}$. 66

3.3 *Profile likelihood —- and adjusted profile likelihood - - - for Kempton example.* 68

5.1 *Field plan for BB trial.* 97

5.2 *Plot of residuals against row and column position with scatterplot smoother.* 100

5.3 *Conditional plots of partial residuals versus column for different row positions.* 101

5.4 *Plot of additive and two-dimensional trend for trial BB.* 106

5.5 *Plot of cultivar means for different spans.* 109

5.6 *Conditional plots residuals, after fitting model (5.1) versus column position for different row positions.* 109
5.7 *Plot of two-dimensional loess with span 28/160 for trial BB without the check plots.* 110
5.8 *Field plan of Block 1 for BJ trial.* 114
5.9 *Plot of partial residuals after analysis of variance of the sub-plot mean yields against sub-plot position. The numbers in brackets are the column numbers in each subplot.* 115
5.10 *Plot of residuals against column position with scatterplot smoother.* . . . 115
5.11 *Conditional plots of partial residuals versus column for different row positions for trial BJ.* 117
5.12 *Plot of estimated trend for row and column position.* 119
5.13 *Two-dimensional surface for unit yields of BJ trial.* 119
5.14 *Plot of the fitted loess smoother for column position and two-dimensional smooth surface.* 123
5.15 *Two-dimensional loess for yields of BJ trials after adjusting for competition.* 125

List of Tables

2.1 *Sample size, n, number of regressors, p, preferred smooth degrees of freedom, d^*, (figure in () is degrees of freedom for rows only), mean percentage error $u_j(d^*)$, and maximum percentage error, $v_j(d^*)$, in standard error of difference for variety effects.* 38

3.1 *Estimated monoculture means for various values of ρ.* 68

4.1 *Values of $\hat{\rho}$, se($\hat{\rho}$), $\hat{\sigma}^2$, and degrees of freedom for the smoother for Jenkyn data* 94

5.1 *Comparison of criteria for selection of the span, in a two-dimensional semi-parametric additive model for trial BB.* 105

5.2 *Anova table for the significance of the treatment effects for BB trial with all plots. C= cultivar, F= fungicide effect. The two-dimensional smooth term lo(row,col) has span 36/320.* 107

5.3 *Treatment means and standard errors for each model.* 108

5.4 *Anova table for the significance of treatment effects of BB trial without the check-plots. The two-dimensional smooth term has span 38/160.* . 110

5.5 *Analysis of variance of the yields of trial BB, with (top) and without (bottom) check plots.* 111

5.6 *Analysis of variance of the half-unit yields of trial BJ.* 116

5.7 *Anova table for the significance of the treatment effects for the unit yields of BJ trial. The two-dimensional smooth term lo(row,col) has span 36/320.* 118

5.8 *Analysis of covariance of the half-unit yields of trial BJ. Column= quadratic term for column position and Row=linear term for row position.* 121

5.9 *Estimates and standard errors of the competition parameter, $\hat{\rho}$, for plot interference (top) and treatment interference (bottom) at the unit level.* 122

5.10 *Estimates and standard errors for the competition parameter for a model with competition between neighbouring half-units.* 124

Acknowledgements

I am immensely grateful to my supervisors Christine Hackett and Iain Currie for their help and constructive criticism, for their encouragement and patience, and for sharing their knowledge of Statistics with me. I am also very grateful to Jim McNicol for his advice and for pointing me in the right direction on many occasions. I also would like to thank Adrian Newton and David Guy for their help and advice with the variety trials. I am grateful to the Scottish Office, Agricultural, Environmental and Fisheries Department for supporting this research. I would like to give especial thanks to Nicole Augustin for her encouragement and many stimulating discussions, and to my parents from whom I have learned the important things in life.

To all of you: Gracias

Abstract

The aim of this project was to develop a joint approach to the estimation of spatial trends and competition effects in agricultural field trials.

We chose to model the trend by means of a semiparametric model and to extend this class of models to include any number of smooth terms. Explicit expressions for the linear and smooth parts of the model are derived. Two approximations to the standard errors of the linear part are presented and compared. We discuss graphical methods for the initial identification of spatial structure in the data and propose more formal procedures to select the degree of smoothing and to test for the significance of treatment effects.

We review the methodology already developed for competition models and improve the fitting procedure by calculating exact adjustments of the profile likelihood for a class of normal regression models. Classical competition models are extended to allow for the estimation of spatial trends via one or more linear smoothers. Methods to estimate the smoothing parameter in the presence of competition were derived. However, we have established that this approach needs to be extended to include correlated errors before it is complete. A mixed model approach to competition was also investigated.

The analysis of the data from two agricultural trials grown at SCRI indicated that SAMs provide a flexible framework for identifying underlying trends in field trials. They generally improve precision of the treatment estimates and they enable spatial trends to be easily visualised. Competition between neighbouring plots was also identified.

Chapter 1

Spatial models for the analysis of agricultural trials: an overview

Large field experiments are often affected by spatial trends across the site and this may bias the estimation of treatment effects and their standard errors. This problem has been approached from two points of view: design and analysis, and in section 1.1 we give an overview of the evolution of the different designs. In section 1.2 we review different methods of analysis. We present in more detail the smoothing approach in section 1.3. Section 1.4 gives a summary of the most relevant work developed in the context of inter-plot competition.

1.1 Design of experiments

Traditional designs, such as randomised blocks (Yates, 1936), attempt to reduce the spatial heterogeneity of agricultural trials by introducing an additive block effect. The underlying assumption is that it is possible to divide the field into experimental units (blocks) in such a way that the environmental variation between plots in the same unit is small compared with the variation between units. However, in an agricultural field trial, the number of cultivars is often large, consequently, blocks are also large and fertility patterns are likely to be present within blocks. A possible solution was to reduce the block size, use incomplete block designs, and recover inter-block information (Nelder, 1969; Patterson and Thompson, 1971), but even with small blocks,

the position of the plots in the field was ignored and randomisation was the only tool to account for the spatial heterogeneity.

In field experiments, the efficiency of parameter estimation will depend on the error variation and on the physical position of the plots in the field, and so over the years, more sophisticated designs such as Latin square designs (Freeman, 1979a; Bailey, 1984 give an extensive review), neighbour-balanced designs (every pair of treatments is on two neighbouring plots an equal number of times, see for example Williams, 1952; Freeman, 1979b) or Alpha designs (Patterson and Williams, 1976; Williams, 1986) have been developed to include the spatial effect into the design. More recently, the introduction of models which incorporate a correlation structure for the errors has led to increased interest in design of experiments for correlated data (Russel and Eccleston, 1987a,b), even in a two-dimensional layout (Eccleston and Chan, 1998).

Most of the designs mentioned above are used for experiments with a single treatment, such as variety trials in which large number of genotypes are compared. If an experiment aims to assess the effect of more than one treatment factor, few of these designs are appropriate. Williams and Jones (1996) proposed a row-column design for factorial experiments, but they point out the difficulties in extending it to more than two treatment factors. However, as Pearce (1983) pointed out *"row-column designs assume that the fertility of any plot can be derived from adding a parameter for rows to one for columns. In general that may work well enough, but a diagonal strip of high or low fertility or a large patch represents so different a pattern as almost to be disastrous"*. Moreover, additional environmental trends can develop in the course of the experiment, such as a gradient in wind-borne disease or patchy infections, and even sophisticated designs might not be able to account for the variation induced by these trends.

1.2 Spatial Models

Spatial models are not an alternative to the analysis of designed experiments but a complement. They aim to produce more accurate estimates of treatment contrasts by removing trends due to different environmental factors. By *spatial or neighbour models* we understand models based on a *smooth trend + error* decomposition (Wilkinson et al., 1983). Spatial models differ from each other in the way the trend is represented. In the next sections we give an overview of these models.

1.2.1 Papadakis analysis of covariance

Papadakis (1937) suggested a model which would become the first alternative to the sophisticated designs that were being developed in the early 1930's (Yates, 1936). Papadakis' model adjusted for local trend by analysis of covariance with respect to the treatment corrected yields of the neighbouring plots (for plots in a single row):

i) Corrected yields for treatment effects give the residuals r_i

$$r_i = y_i - \text{mean of plots with treatment } [i]$$

treatment $[i]$ refers to treatment in plot i

ii)
$$z_i = \begin{cases} (r_{i-1} + r_{i+1})/2 & 2 \leq i \leq n-1 \\ r_2 & i = 1 \\ r_{n-1} & i = n \end{cases}$$

iii) The resulting variable is used in the analysis of covariance

$$y = \mu + \tau_{[i]} + \beta z_i + \epsilon_i$$

where $\tau_{[i]}$ is the treatment allocated to plot i

This method of analysis has been discussed by several authors (Bartlett, 1938; Binns and Jui, 1985; Zimmerman and Harville, 1989, etc). Bartlett (1978) extended Papadakis' model to plots arranged in a rectangular grid and suggested that the method might be improved by an iterative version of Papadakis' model. This method is more efficient than the randomised block design when the number of treatments is large and a trend component is not included in the model. However, it has proved to be

inefficient when there are substantial trend effects in the data (Wilkinson et al., 1983).

In the same context of analysis of covariance, Wiancko (1914) introduced the method of *check plots*. The method uses the yield from neighbouring plots to calculate a fertility index for each plot. Check plots are arranged systematically throughout the trial and a standard variety is used for the check plots. The method of check plots has been used mostly in cases where there are a large number of cultivars and replication is limited. In chapter 5 we give an example of this analysis.

Wilkinson et al. (1983) introduced a model inspired by Papadakis', but the nearest-neighbour covariate is not corrected for treatment effects and blocks are incorporated in the model. The method was regarded as "*moving block*", an extension of the classic fixed block analysis. Wilkinson et al. (1983) proposed the analysis of agricultural trials based on the "trend plus error" decomposition on which most spatial models have been based. Some analyses assume that the trend is fixed while others see it as the realisation of a stochastic process. Least squares smoothing methods (Green et al., 1985) are a well known example of spatial models under the assumption of a fixed smooth trend. These methods are closely related to the general smoothing formulation and we will discuss them in detail later in the chapter. First, we give an overview of spatial models in which the trend is regarded as random.

1.2.2 Trend as a random process

Models assuming a random trend use different approaches to account for spatial correlation: differencing, using partial residuals, etc. Many authors justify this interpretation of the trend by the gain in accuracy of treatment estimates compared with sophisticated designs (Besag, 1984). However, they fail to give an intuitive representation of the biological process occurring in the field.

First difference model

Besag and Kempton (1986) introduced two methods of analysis based on a stochastic

description of plot yields. The random part of the models is the combination of a fertility model plus a random error. The *first differences* model may be written as

$$Y = \gamma 1 + T\tau + X.$$

Therefore Y is represented by a fixed treatment effect, τ, plus a stochastic fertility process X. The estimation of the treatment effects is then based on the vector of first differences $U = \Delta y$. An extension of this model is the model with *errors in variables*; the formulation of the model is similar to the first differences model, but a random error is also superimposed allowing for comparison between the experimental and fertility variation. The method of intra-Nearest Neighbours of Wilkinson et al. (1983) and the iterated method of Papadakis are particular cases of this model.

Time series approach

There is an extensive literature on models which represent the trend using a time series approach. Gleeson and Cullis (1987) proposed that the trend could be represented as an autoregressive integrated moving average (ARIMA) process, and they used the residual maximum likelihood of Cooper and Thompson (1977) to estimate the variance parameters. Gleeson and Cullis (1987) write the model as

$$y = D\tau + \psi + \eta,$$

where D is the design matrix, ψ is the random trend represented by an ARIMA(p,d,q) model and η is the vector of random errors.

Cullis and Gleeson (1991) extended this model by introducing a two-dimensional component; in this case, the errors are modelled by a subclass of separable lattice process (if $\rho(g,0)$ and $\rho(0,h)$ are the lag-g and lag-h correlations of a one-dimensional processes for rows and columns, separability means $\rho(g,h) = \rho(g,0)\rho(0,h)$). Several authors (Lill et al., 1988; Martin, 1990; Kempton et al., 1994; Brownie and Gumpertz, 1997) have discussed the accuracy and efficiency of these methods and showed that these models improved the row + column analysis. Gilmour et al. (1997) extended the model proposed by Cullis and Gleeson (1991) by identifying three different sources of spatial variation: a non-stationary large scale variation across the field (global trend),

stationary variation within the trial (local trend, following Zimmerman and Harville (1991)) and extraneous variation due to experimental procedures. They proposed the use of polynomials or cubic smoothing splines (Verbyla et al., 1998) to represent the global trend (as an alternative to differencing), and a separable $AR \times AR$ variance model for the plot errors. The extraneous variation is modelled through fixed or random row and column effects. One of the main difficulties in applying these models is the selection of the appropriate ARIMA process. Martin (1990) and Cullis and Gleeson (1991) suggested the use of the spatial correlation matrix of the residuals. In the same context, Tunnicliffe Wilson (1989) used the marginal likelihood for model selection. Gilmour et al. (1997) and Verbyla et al. (1998) used the sample variogram of the residuals and likelihood ratio tests to identify the appropriate model. These models have been recently applied to the analysis of multi-environment early generation variety trials (Gogel et al., 1995) and longitudinal data (Verbyla et al., 1998).

Random fields

The random field approach of Zimmerman and Harville (1991) is closely related to geostatistical kriging analysis. The novel idea introduced by Zimmerman and Harville (1991) was the decomposition of the spatial variation into a large-scale dependence (or trend), estimated through the mean, and a small-scale dependence estimated via the correlation structure. (This decomposition of the spatial heterogeneity has also been a recent subject of research in Verbyla et al., 1998). In this model, the yields are assumed to be the realisation of a random field. Zimmerman and Harville (1991) used polynomials to account for the large scale variation, suggested several covariance functions (spherical, exponential, Gaussian) and proposed the use of cross-validation to chose the appropriate covariance function. Zimmerman and Harville (1991) also showed how most spatial models are closely related to each other and can be seen as applications of generalised least squares to a general model

$$Gy = A\alpha + GT\tau + \epsilon,$$

where α and τ are vectors of unknown parameters, T is the design matrix and G is a matrix of known constants (i.e. the row and column incidence matrices in a row and

column analysis) or functions of y (as in Papadakis' analysis of covariance).

1.3 Non-parametric regression

We propose the use of nonparametric smoothing methods to account for the spatial variation in agricultural field trials. This might be seen as the nonparametric extension of the method introduced by Papadakis (1937). The idea is to model the trend as a function of the plot position in the field, and model the regression function nonparametrically letting the data indicate the functional form.

Non-parametric regression is an approach to fitting curves and surfaces to data by smoothing and is a natural extension of ordinary linear regression with the advantage that the function f does not have a rigid form chosen in advance: the shape of the curve will be determined by the data. We give a summary of some of the different smoothing methods (Simonoff, 1996; Cleveland and Loader, 1996; Bowman and Azzalini, 1997 give a detailed description) in the case of Gaussian errors, and describe additive and semiparametric models in more detail.

The model underlying a smoothing regression problem is

$$y = f(t) + \epsilon, \tag{1.1}$$

where f is a smooth function of the covariate t and the errors are assumed to be uncorrelated (generally Gaussian). An estimate of f, $\hat{f}$, is called a *smoother* and a smoother is **linear** if

$$\hat{f} = Sy,$$

and S is called the smoother matrix. Among the properties of the smoother matrix of most linear smoothers is the fact that it has two eigenvalues equal to 1, since it satisfies $S1 = 1$ (it reproduces constants) and $St = t$ (it leaves linear functions of the predictor invariant). Linear smoothers fall into two groups: those specified by the fitting procedure (kernel smoothers) and those which are the solution to a

minimisation problem (splines). Smoothers in the first group are easy to understand and interpret, while the ones in the second group have more attractive properties.

1.3.1 Kernel Methods

Most smoothers do some sort of local averaging within a neighbourhood of the target point. The neighbourhood may be symmetric or non-symmetric (the nearest neighbours to the target point are chosen regardless of which side of the target point they lie). The size of the neighbourhood is determined by a smoothing parameter. The smoothers differ in the way the averaging is performed within the neighbourhood. A kernel smoother uses a set of local weights to produce the fitted value at each target point, and the weights usually decrease as the distance from the target point increases. One of the simplest examples is the **running mean** smoother, where the weights in the neighbourhood are constant and equal to $1/\lambda$ (λ being the size of the neighbourhood). Other kernel smoothers use more complicated weighting functions (Nadayara, 1964; Gasser and Müller, 1979, 1984). An important issue is the behaviour of the smoother at the endpoints; several authors (Müller, 1987; Chu and Marron, 1991) have studied different kernel smoothers in detail and reported bias problems at the endpoints, for example, smoothers which use equal weights for all points in a symmetric neighbourhood will tend to flatten out trends near the end points and can be biased. A class of smoothers which perform satisfactorily at the end points are the local polynomial estimators. The fitted value at each target point is calculated by fitting a polynomial by weighted least squares (in the case of Gaussian errors) or local-likelihood (Hastie and Tibshirani, 1986, for other error distributions).

Locally-weighted running line

We describe in more detail the locally-weighted running line smoother or **loess** of Cleveland (1979) (we use this smoother and its two-dimensional version in the analysis of field trials). Given model (1.1), the fitted value $\hat{y}_i$ at each point t_i is the fitted value of a polynomial (in this case a line) estimated using weighted least squares. The weights w_j at each point t_j in the neighbourhood of t_i are calculated using the

tri-cube weight function:

$$w_j = \begin{cases} (1 - u(t_j)^3)^3 & \text{for } 0 \leq u(t_j) < 1 \\ 0 & \text{otherwise} \end{cases}$$

where $u(t_j) = |t_i - t_j|/\Delta(t_i)$ and $\Delta(t_i)$ is the distance of the furthest point in the neighbourhood from t_i. Then, $\hat{y}_i$ is obtained by minimising

$$\sum_{j=1}^{n} w_j(y_j - \alpha - \beta(t_j - t_i))^2.$$

In the case of a two-dimensional loess, each target point t_i has two components, $t_i = (u_i, v_i)$. The neighbourhood is calculated using the Euclidean distance and the weights are defined as above. Then we find values of α, β and γ that minimise

$$\sum_{j=1}^{n} w_j(y_j - \alpha - \beta(u_j - u_i) - \gamma(v_j - v_i))^2.$$

If a quadratic polynomial is used, the terms, u, v, uv, u^2 and v^2 are fitted locally.

Devlin (1986) and Cleveland et al. (1988) discussed the properties of local polynomial estimators (Hastie and Loader, 1993; Cleveland and Loader, 1996 gave a general description of local regression methods) and Cleveland and Loader (1996) derived exact expressions for the bias and variance of the loess smoother. Cleveland and Devlin (1988) showed some applications of loess to data exploration and diagnostic checking, extended the univariate locally weighted smoother to a two-dimensional surface and gave details on the distribution of residuals. They also proposed an extension of the ordinary F test used in analysis of variance to compare two models with different degrees of smoothing.

1.3.2 Smoothing splines

The roughness penalty approach to smoothing might be seen as a compromise between linear regression models and non-parametric regression methods. Authors such as Silverman (1985) and more recently Green and Silverman (1994) gave an extensive overview of the spline smoothing approach. A cubic smoothing spline may be seen

as the curve $f(t)$ which minimises the following penalised residual sum of squares,

$$\sum_{i=1}^{n}(y_i - f(t_i))^2 + \lambda \int (f''(t))^2 dt. \tag{1.2}$$

Then, $\hat{f} = (I + \lambda K)^{-1} y$ (where K satisfies $\int (f''(t))^2 dt = f' K f$; see for example Green and Silverman, 1994 for a definition of K). Eubank (1984) studied the properties of the smoother matrix associated with smoothing splines (we give more detail in chapter 4). The penalising constant, λ, has the same role as the span in the locally-weighted running line and controls the smoothness of the curve: the larger λ, the smoother the curve. Although a smoothing spline is the curve that minimises a certain penalised criterion, it may also be seen as a kernel smoother. Silverman (1984) gave an expression for the equivalent kernel of a smoothing spline, showing that the bandwidth depends on the smoothing parameter, the sample size and the density of the design points and Müller (1987) demonstrated that for an appropriate choice of the smoothing parameter, loess and cubic smoothing splines are asymptotically equivalent. Splines also admit a Bayesian characterisation (Wahba, 1978): if f has the following improper prior distribution, $f \sim N(0, K^{-}\tau^2)$ (where K^{-} is the generalised inverse of the matrix defined above), then, the smoothing spline is the mean of the posterior distribution of f, $f|y \sim N(S(\lambda)y, S(\lambda)\sigma^2)$, where λ is the smoothing parameter, $\lambda = \sigma^2/\tau^2$. A natural extension of the smoothing spline to two dimensions is the *thin plate spline*. The penalty function is a combination of derivatives of the surface with respect to the two covariates. Thin plate splines have many applications (Green and Silverman, 1994), however their computational implementation is more complex than the univariate case and they are not as widely used as loess.

1.3.3 Further topics in smoothing

Degrees of freedom

There are several definitions of the degrees of freedom fitted to the data in model (1.1), depending on the context in which they are used (Buja et al., 1989). If we make a comparison with the regression model, the number of degrees of freedom would be the number of parameters that are fitted to a particular value of the smoothing

parameter. Green et al. (1985) defined the *equivalent degrees of freedom* as the trace of the smoother matrix $df = \text{tr}(S)$. In the context of model comparison, the degrees of freedom associated with the residual sum of squares would be $\text{tr}(I - S)'(I - S) = n - \text{tr}(2S - S'S)$, and therefore another possible definition for the degrees of freedom would be $df = \text{tr}(2S - S'S)$. Finally, in the linear model, $E(Var(\hat{y})) = p\sigma^2$, where $p = degrees\ of\ freedom$; in the smoothing context, the analogous definition would be $df = \text{tr}(SS')$. When the smoother is symmetric with eigenvalues in [0,1], the three definitions are related as follows,

$$\text{tr}(SS') \leq \text{tr}(S) \leq \text{tr}(2S - S'S).$$

The most common definition and the easiest to calculate is $\text{tr}(S)$. The definition of the residual degrees of freedom will depend on which definition we use for the degrees of freedom associated with the smoother (Buckley et al., 1988 gave a review of the different definitions). Hastie and Tibshirani (1990) also gave an approximation for $\text{tr}(2S - SS')$ based on $\text{tr}(S)$: $\text{tr}(2S - SS') \approx 1.25\text{tr}(S) - 0.5$.

Smoothing parameter selection

Much research has been done on how best to choose the smoothing parameter and many methods have been developed. However, which method is the best is still a matter of debate. The smoothing parameter controls the shape of the smooth curve. A large neighborhood produces a smooth curve with low variance but possibly biased. This is known as the *bias-variance trade-off*. A reasonable way to choose the smoothing parameter is to use a criterion that combines the bias and the variance: the mean integrated squared error or its discrete analogue, the average mean square error. The cross-validation criterion (Stone, 1974) minimises an estimate of the mean squared error over a range of values for the smoothing parameter.

$$CV = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{f}_\lambda^{-i})^2,$$

where $\hat{f}_\lambda^{-i}$ is the estimate of f with the $i - th$ data point omitted at a particular value of the smoothing parameter λ. Hastie and Tibshirani (1990) gave an explicit expression for cross-validation (CV) criterion in the case of linear smoothers. The

generalised-cross-validation criterion, GCV, is a version of the CV criterion which is easier to compute.

Another criterion is the Akaike information criterion (Akaike, 1973), AIC, which minimises the Kullback-Leiber discrepancy (a likelihood-based method). More recently, Hurvich et al. (1998) and Simonoff and Tsai (1999) proposed a modified version of the AIC criterion which avoids the tendency to under-smooth seen in the methods described above. We give more detail on these methods in chapter 5.

In the context of density estimation, the so-called plug-in selectors (for example the one given in Ruppert et al., 1995) are very popular. These methods write the bias of $\hat{f}$ as a function of the unknown f. An estimate of f is then plugged-in to derive an estimate of the mean integrated squared error. Loader (1995) gave a critical review of these methods and compared them with more classical methods.

All the selection criteria presented above assume that the errors are uncorrelated. When the errors are correlated, it is well known that methods such as CV, GCV, etc under-smooth the data. Altman (1990) proposed two methods of estimating the smoothing parameter in the presence of correlation. More recently, Wang (1998b) has developed a modified cross-validation criterion (for smoothing splines) which allows for the simultaneous estimation of the smoothing and correlation parameters. A similar approach is taken by Verbyla et al. (1998), but they used REML for the parameter estimation.

1.3.4 Additive models

Additive models are an extension of the multiple linear regression models to models which incorporate several non-parametric smooth terms,

$$y = \alpha + \sum_j f_j(t_j) + \epsilon, \tag{1.3}$$

where f_j are smooth functions estimated by one or two-dimensional linear smoothers (splines, loess, etc), ϵ is a vector of independent, normally distributed errors with

$E(\epsilon) = 0$ and $Var(\epsilon) = \sigma^2 I$, and $E(y) = \alpha$ (the additive model is not identifiable unless this condition is imposed). Buja et al. (1989) discussed additive models in detail; they give an extensive review of linear smoothers and proposed the backfitting algorithm, and an improved version of it, as a method for fitting these models (see chapter 2 for more details). Convergence properties are also studied in detail, especially for the two smoother case, for which they give explicit expressions for the solutions. The convergence of the backfitting algorithm depends among other things on the degree of correlation between the predictors, and therefore the more additive terms we have in the model, the more difficult it is to reach convergence. Buja et al. (1989) discussed the degrees of freedom associated with an additive model. The calculation of the trace of the matrices involved is quite complicated, so they suggested approximating the degrees of freedom by the sum of the degrees of freedom associated with each individual smoother (as in the linear case). More recently, Opsomer (1997) studied the bivariate additive model for locally polynomial smoothers and Durban et al. (1999) gave a closed form for the hat matrix of an additive model with any number of smoothers, extending the results given in Buja et al. (1989). There are other approaches to the estimation of additive models. Hastie and Tibshirani (1990) and Green and Silverman (1994) gave the additive version of the penalised least squares approach in the case of cubic smoothing splines. Hastie and Tibshirani (1990) also gave a Bayesian version of additive models and Hastie and Tibshirani (1998) have recently introduced a Bayesian backfitting algorithm. Hastie and Tibshirani (1986, 1987, 1990) extended additive models to a more general setting, allowing for non-Gaussian errors.

Additive models are a flexible method for identifying non-linear covariate effects and have been applied as a data-driven method in many disciplines (see for example Hastie and Tibshirani, 1987, 1990).

1.3.5 Semiparametric models

Semiparametric models are traditionally defined as models in which all but one term are assumed to be linear,

$$y = X\beta + f(t) + \epsilon,$$

where X is the design matrix, and the vector of regression coefficients β and the unknown curve $f(t)$ are to be estimated. Schimeck (1997) and Green and Silverman (1994) gave a summary of different approaches to estimation in semiparametric models.

Green et al. (1985) introduced the method of *Least Squares Smoothing* in the context of agricultural trials, assuming the Wilkinson et al. (1983) decomposition of the environmental effects into a smooth trend plus independent error. The term $f(t)$ is regarded as smooth if it is locally linear, that is, $f(t_{i-1}) - 2f(t_i) + f(t_{i+1}) \approx 0$, or in matrix notation, $\Delta f \approx 0$ (Δ is the matrix of second differences). Then, Green et al. (1985) estimated β and f by solving the following set of equations

$$\begin{aligned} f &= \mathcal{F}(y - X\beta) \\ \beta &= \mathcal{H}(y - f), \end{aligned} \tag{1.4}$$

where $\mathcal{H}$ is an estimator of the treatment effects (for example $(X'X)^{-1}X'$) and $\mathcal{F}$ is a linear smoother. This approach is also used in Buja et al. (1989), as a particular case of an additive model with two smoothers where the smoother corresponding to the linear part of the model is the projection smoother $X(X'X)^{-1}X'$.

Green (1987) examined penalised likelihood in the context of semiparametric models, where β and f are chosen to minimise,

$$\sum_i (y_i - X_i'\beta - f(t_i))^2 + \lambda \int f''(t)^2 dt.$$

Green (1987) showed that this method is equivalent to least squares smoothing when the errors are normally distributed. A similar approach is taken by Heckman (1986) who studied the asymptotic distribution of the regression terms. Also in the context of field trials, Green (1985) demonstrated that generalised least squares analysis of a linear model is a smoothing method and that different variance matrices result in

different models. Green (1985) showed that least squares smoothing is equivalent to a generalised least squares regression of Δy on ΔX with $\sigma^2(\lambda^{-1}I + \Delta\Delta')$ as variance matrix for Δy (for Δ defined as above).

The methods mentioned above use the same amount of smoothing for the parametric and nonparametric part of the model. Rice (1986) showed that the penalised least squares estimates of β may be biased and the bias may be reduced by undersmoothing the nonparametric component, suggesting that two different smoothing parameters should be used. The method of Speckman (1988) is inspired by this principle. He discussed an alternative to least squares smoothing motivated by partial residual analysis. The estimate of β is the ordinary least squares estimate based on the adjusted response, $(I - S)y$ and regression matrix $(I - S)X$ (where S is the smoother matrix of any linear smoother). The method is equivalent to penalised least squares if the S is the smoother matrix of a projection smoother. Speckman (1988) showed that the estimate of β obtained by this method had less bias. Cuzick (1992) also studied and compared the asymptotic properties of this estimate. However, as Green and Silverman (1994) pointed out *"more practical experience is required before it is possible to see whether the apparent theoretical advantage of the Speckman approach offsets the loss of interpretability and flexibility in departing from the penalised least squares paradigm"*.

Severini and Wong (1992) proposed a general approach to estimating the parametric component in a semiparametric model: the nonparametric component is considered as a nuisance parameter and profile likelihood is used to estimate β. Kauermann (1997) used the modified profile likelihood of McCullagh and Tibshirani (1990) and obtained, in the Gaussian case, similar estimates of β to Speckman (1988).

Recently, Wang (1998a) and Verbyla et al. (1998) gave a mixed model representation of cubic smoothing splines and used the mixed model equations of Henderson et al. (1959) to estimate β and $f(t)$. Conditional on the smoothing parameter, the solutions of those equations coincide with the least squares smoothing estimates of

Green et al. (1985).

1.4 Competition Models

Another way in which neighbouring plots may be related to each other is through inter-plot competition. When varieties are grown in small plots with different cultivar neighbours, systematic bias may appear. Several authors (Federer and Basford, 1991; Azaïs et al., 1993), studied neighbour-balanced designs in which every pair of treatments is on two neighbouring plots an equal number of times. This ensures that treatment comparisons are as little affected by competition effects as possible. However, these designs require an amount of replication which is not always possible, especially when there are a large number of cultivars. Kempton (1982) proposed a method for correcting for competition effects in yield trials through a common competition effect, ρ, acting on a covariate, equal to the mean yield of the neighbouring plots. (We study this model in more detail in chapters 3 and 4). Kempton (1982) showed that the least squares estimate of ρ is incorrect and proposed the use of maximum likelihood to estimate it. A model for varieties with different competition coefficients is also studied in that paper. In this model, the competition model for variety r with variety s as a neighbour is decomposed into $\rho_{rs} = \delta_r \gamma_s$, where δ_r is a measure of the sensitivity of variety r to competition and γ_s is a measure of the of aggressiveness of variety s. Kempton (1985) discussed this method in more detail and Talbot et al. (1995) applied it to several winter wheat trials.

Besag and Kempton (1986) gave a simultaneous autoregressive formulation for the original model given in Kempton (1982) and extended it to include block effects,

$$y = X\beta + T\gamma + \rho W y + \epsilon,$$

where β and γ correspond to treatment and block effects, and W is the neighbouring matrix. Besag and Kempton (1986) pointed out that one underlying assumption in this model is that the fertility effect is much smaller that the inter-plot competition. Connolly (1992) proposed an efficient numerical method based on the profile likeli-

hood for competition effects and investigated the adequacy of the standard error of the competition parameter. This method has been used to analyse root crops (Besag and Kempton, 1986; Connolly et al., 1993).

In crops such as beans or cereal, the competitive effect may be associated with varietal differences in height, rather than varietal yields. Several authors (Pearce, 1957; Draper and Guttman, 1980) considered models of this type in which each treatment is supposed to have an effect on the plot to which it is applied and also on the neighbouring plots. Besag and Kempton (1986) gave a general formulation for the model for interference between neighbouring treatments,

$$y = B\eta + T\tau + RT\theta + \epsilon, \tag{1.5}$$

where B and T are the design matrices for block and treatment effects, R is the neighbouring incidence matrix, θ is the vector of centred treatment neighbour effects and ϵ is the vector of independent errors. Draper and Guttman (1980) gave a simpler version of this model in which $\theta = \rho\tau$ and the number of parameters fitted is much smaller. Kempton and Lockwood (1984) analysed several variety trials of field beans using model (1.5) and analysis of covariance using height difference of neighbours as a covariate.

When strong fertility trends are present in the field and block sizes are large, the effect of the trend will mask the effect of competition and alter the magnitude of the competition parameter and it might be difficult to separate both effect. None of the competition models developed until now include a term to model fertility trends. The work we present here aims to give a possible solution to this problem.

Chapter 2

Semiparametric additive models

2.1 Introduction

Semiparametric models with a single smooth term (Green et al., 1985) have been used to analyse data from field trials when plots are arranged in a single row, using the plot's position as the covariate for the smooth part of the model. However, large field experiments are likely to be arranged in a rectangular grid; in this case, yield (or any other crop measurement) may be modelled as the sum of smooth functions of position in the field (row and column position) or as a two-dimensional smooth surface. Other smooth terms, such as disease score or plant height, may also be included. Therefore, it is necessary to extend semiparametric models to models which combine parametrised treatment effects with any number of smooth terms and derive explicit expressions for the parametric part of the model and the standard errors. We call this model a *semiparametric additive model* (SAM).

We will interpret a SAM as a particular case of an additive model (Buja et al., 1989) in which the smoother corresponding to the linear part of the model is a projection smoother. There are several approaches to the estimation of additive models; the most general method of estimation is the *backfitting algorithm* which solves the estimating equations (which are a *smooth* version of the normal equations in multiple linear regression). This algorithm is computationally very efficient but, as an

iterative procedure, it does not give explicit expressions for the estimated smooth terms and, in particular, for the estimates of the parameters in the linear part of the model. As an alternative, we give explicit solutions to the normal equations and the conditions for uniqueness in Section 2.2, together with some properties of the hat matrix of an additive model. In Section 2.3 we concentrate on inference in SAMs. First, we give expressions for the estimates of the linear parameters and second we extend the non-parametric method introduced by Speckman (1988) to the case of multiple smooth terms. An exact and an approximate method for the calculation of the standard errors are presented in Section 2.4 and are illustrated with several examples of agricultural field experiments.

2.2 Additive models

A simple smoothing regression model has the form

$$y = f(z) + \epsilon$$

and f is usually estimated by a linear smoother

$$\hat{f} = S^* y,$$

where S^* is a matrix called a *smoother matrix*. We define the *centred smoother matrix* (see Hastie and Tibshirani, 1990, pp 114-115) associated with f as

$$S = (I - 11'/n)S^*. \tag{2.1}$$

The smoother matrix of a linear smoother satisfies

$$S^*1 = 1 \tag{2.2}$$

$$S1 = 0 \quad \text{from (2.2) and (2.1)} \tag{2.3}$$

$$1'S = 0' \quad \text{from (2.1).} \tag{2.4}$$

We will say that a smoother is centred if its associated smoother matrix is centred.

In an additive model (Buja et al., 1989), we suppose a variable y can be modelled by

the sum of q smooth terms which act on explanatory variables $z_1, ..., z_q$. Thus, with observations $y_1, ..., y_n$ and q smooth functions f_i, we assume

$$y = \alpha + f_1 + f_2 + ... + f_q + \epsilon. \tag{2.5}$$

We will assume that $E(f_i) = 0$ to avoid problems with the identifiability of the model and, therefore, $\alpha = E(y)$ will be estimated by the average of the observed values, $\hat{\alpha} = \bar{y} = (11'/n)y$. Further, we will assume that S_j, the smoother matrix associated with f_j, is centred.

If the additive model (2.5) is correct,

$$f_j(z_j) = E(y - \alpha - \sum_{k \neq j} f_k(z_k)|z_j) = P_j(y - \alpha - \sum_{k \neq j} f_k(z_k)), \quad j = 1, \ldots, q$$

where P_j is the conditional expectation operator $E(.|z_j)$. Equivalently, we may find f_i by solving the following set of linear equations whose data version (substituting P_j by S_j) is:

$$\begin{pmatrix} I & S_1 & S_1 & ... & S_1 \\ S_2 & I & S_2 & ... & S_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ S_q & S_q & S_q & ... & I \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_q \end{pmatrix} = \begin{pmatrix} S_1 \\ S_2 \\ \vdots \\ S_q \end{pmatrix} (y - \hat{\alpha}), \tag{2.6}$$

where S_j is the centred smoother matrix associated with f_j and defined in (2.1). The set of equations given in (2.6) are called the *normal equations* for the additive model (2.5) and are similar to the normal equations of a multiple regression model.

2.2.1 Solution to the normal equations

The system of estimating equations (2.6) is an $nq \times nq$ system and, therefore, unless n is very small, it is computationally very expensive to solve the equations using a standard procedure. An alternative method is the **backfitting algorithm**, which is an iterative procedure given by:

1. Initialise: $f_i = f_i^{(0)}, i = 1, \ldots, q$

2. Cycle $i = 0, ..., q, 0, ..., q, ...$

$$f_i = S_i(y - \bar{y} - \sum_{\substack{j=1 \\ j \neq i}}^{q} f_j), \ i = 1, ...q.$$

3. Continue until the individual functions f_i converge.

The backfitting algorithm estimates the fitted smooth values efficiently but, as an iterative procedure, it does not give an explicit expression for the estimate of the individual smooth functions or for the hat matrix of the fitted model. From the point of view of semiparametric additive models, it is of interest to obtain a closed form for the estimate of the regression part of the model, and therefore, it is necessary to solve the system of estimating equations directly.

The identifiability condition $E(f_i) = 0$ together with $\hat{\alpha} = \bar{y}$ means that at each step of the backfitting algorithm, $\hat{f}_i$ has mean zero and therefore it is necessary to use a *centred smoother matrix.*

We define M_q^* as the hat matrix for model (2.5) when the backfitting algorithm is used to estimate f_i. Further, each of the sub-models $y = \alpha + f_1 + f_2 + \ldots + f_i + \epsilon$ gives rise to a hat matrix M_i^*, $i = 1, \ldots, q-1$. We define the centred hat matrix M_i as the hat matrix for model (2.5) when the response has been centred. Since,

$$\hat{y} = M_q^* y = \hat{\alpha} + \hat{f}_1 + \ldots + \hat{f}_q = (11'/n)y + M_q y,$$

M_i is given by

$$M_i = M_i^* - 11'/n. \tag{2.7}$$

We will show how the centred hat-matrix M_q for the model $y = \alpha + f_1 + \ldots + f_q + \epsilon$ is related to the centred smoothers used to estimate the f_i in the models $y = \alpha + f_i + \epsilon$, $i = 1, \ldots, q$.

Lemma 2.1 *The hat matrix M_k^* of an additive model with k smooth terms, $y = \alpha + f_1 + \ldots + f_k + \epsilon$ and a centred smooth matrix S_{k+1} as defined in (2.1) satisfy*

$$S_{k+1} M_k = S_{k+1} M_k^* \tag{2.8}$$

$$11'/n(I - S_{k+1} M_k)^{-1} = 11'/n \tag{2.9}$$

Proof From the definition of a centred hat matrix given in (2.7)

$$\begin{aligned} S_{k+1}M_k &= S_{k+1}(M_k^* - 11'/n) \\ &= S_{k+1}M_k^* - S_{k+1}11'/n \\ &= S_{k+1}M_k^* - 0 \quad \text{by (2.4)} \end{aligned}$$

Proof of (2.9) follows immediately from (2.4). ∎

In the next theorem we give an expression for the sequence of centred hat matrices M_i, $i = 1, \ldots, q$.

Theorem 2.1 *Provided the necessary inverses exist, the sequence of centred hat matrices M_i, $i = 1, \ldots, q$, satisfies the recurrence relations*

$$M_1 = S_1; \quad M_i = I - (I - S_i)(I - M_{i-1}S_i)^{-1}(I - M_{i-1}),\ i = 2, ..., q. \tag{2.10}$$

Proof For $i = 1$ the backfitting estimate of f_1 in the model

$$y = \alpha + f_1 + \epsilon$$

is given by

$$\hat{f}_1 = S_1(y - \bar{y}) = S_1(I - 11'/n)y = S_1 y \quad \text{by (2.3)}.$$

Then,

$$\hat{y} = \bar{y} + S_1 y = (11'/n + S_1)y$$

and hence,

$$M_1^* = 11'/n + S_1 \quad \text{and} \quad M_1 = S_1 \quad \text{by (2.7)}.$$

Next, consider the estimating equations for an additive model with $k + 1$ smooth functions $f_1, \ldots, f_{k+1}$ with centred smoother matrices $S_1, ..., S_{k+1}$,

$$y = \alpha + f_1 + \ldots + f_{k+1} + \epsilon. \tag{2.11}$$

The first k equations are

$$\hat{f}_i = S_i(y - \bar{y} - \sum_{\substack{j=1 \\ j \neq i}}^{k+1} \hat{f}_j),\ i = 1, ..., k$$

which we write as

$$\hat{f}_i = S_i(y^* - \bar{y} - \sum_{\substack{j=1 \\ j \neq i}}^{k} \hat{f}_j),\ i = 1, ..., k$$

where $y^* = y - \hat{f}_{k+1}$. The hat-matrix for $y^* = \alpha + f_1 + ... + f_k + \epsilon$ is M_k^*, so

$$\hat{y}^* = \hat{\alpha} + \hat{f}_1 + ... + \hat{f}_k = M_k^* y^*, \tag{2.12}$$

where $\hat{\alpha} = \bar{y} - (11'/n)\hat{f}_{k+1}$ (the mean of the response) and, hence

$$\hat{f}_1 + ... + \hat{f}_k = M_k^* y^* - \bar{y} + (11'/n)\hat{f}_{k+1} \text{ (from the definition of } y^*). \tag{2.13}$$

Now substitute for $\hat{f}_1 + ... + \hat{f}_k$ in the $(k+1)$st estimating equation to get

$$\hat{f}_{k+1} = S_{k+1}(y - (11'/n)\hat{f}_{k+1} - M_k^*(y - \hat{f}_{k+1})).$$

From (2.8) and the fact that $S1 = 0$ we obtain,

$$\hat{f}_{k+1} = S_{k+1}(y - M_k(y - \hat{f}_{k+1}))$$

which yields

$$\hat{f}_{k+1} = \Delta S_{k+1}(I - M_k)y, \tag{2.14}$$

where we have defined

$$\Delta = (I - S_{k+1}M_k)^{-1}. \tag{2.15}$$

From (2.4) and (2.9) , it follows that

$$(11'/n)\hat{f}_{k+1} = 0 \tag{2.16}$$

$$(11'/n)y^* = \bar{y}. \tag{2.17}$$

Thus, (2.13) becomes

$$\begin{aligned} \hat{f}_1 + ... + \hat{f}_k &= M_k^* y^* - \bar{y} \quad \text{by (2.16)} \\ &= M_k(y - \hat{f}_{k+1}) + (11'/n)y^* - \bar{y} \quad \text{by (2.7)} \\ &= M_k(I - \Delta S_{k+1}(I - M_k))y \quad \text{by (2.14) and (2.17)} \\ &= M_k \Delta (I - S_{k+1})y. \end{aligned} \tag{2.18}$$

Using the formula for the inverse of the sum of matrices (see for example, Schott, 1997):

$$(D - CA^{-1}B)^{-1} = D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1}, \tag{2.19}$$

we rewrite (2.15) as

$$\Delta = I + S_{k+1}(I - M_k S_{k+1})^{-1} M_k. \tag{2.20}$$

Substituting (2.20) into (2.14) and (2.18) we find

$$\hat{f}_{k+1} = S_{k+1}(I - M_k S_{k+1})^{-1}(I - M_k)y \tag{2.21}$$

$$\begin{aligned} \hat{f}_1 + \ldots + \hat{f}_k &= [M_k + M_k S_{k+1}(I - M_k S_{k+1})^{-1} M_k](I - S_{k+1})y \\ &= [I - (I - M_k S_{k+1})^{-1}(I - M_k)]y. \end{aligned} \tag{2.22}$$

Therefore,

$$\hat{y} = \hat{\alpha} + \hat{f}_1 + \ldots + \hat{f}_{k+1} \tag{2.23}$$

$$\begin{aligned} &= \bar{y} + [S_{k+1}(I - M_k S_{k+1})^{-1}(I - M_k)]y + [I - (I - M_k S_{k+1})^{-1}(I - M_k)]y \\ &\quad \text{by (2.21) and (2.22)} \\ &= \bar{y} + [I - (I - S_{k+1})(I - M_k S_{k+1})^{-1}(I - M_k)]y. \end{aligned} \tag{2.24}$$

Let M^*_{k+1} be the hat matrix for model (2.11), then $\hat{y} - \bar{y} = (M^*_{k+1} - 11'/n)y$; by the definition of the centred hat matrix given in (2.7), we obtain

$$M_{k+1} = I - (I - S_{k+1})(I - M_k S_{k+1})^{-1}(I - M_k). \tag{2.25}$$

■

Remark 1 From the expressions for $\hat{f}_{k+1}$ and $\hat{f}_1 + \ldots + \hat{f}_k$ given in (2.14) and (2.18) respectively and Δ defined by (2.15) we find

$$\begin{aligned} M_{k+1} &= M_k\Delta(I - S_{k+1}) + \Delta S_{k+1}(I - M_k) \\ &= \Delta(I - S_{k+1}) - (I - M_k)\Delta(I - S_{k+1}) + \Delta S_{k+1}(I - M_k) \\ &= \Delta(I - S_{k+1}M_k) - (I - M_k)\Delta(I - S_{k+1}) \quad . \\ &= I - (I - M_k)\Delta(I - S_{k+1}) \quad \text{by (2.15)} \\ &= I - (I - M_k)(I - S_{k+1}M_k)^{-1}(I - S_{k+1}). \end{aligned} \tag{2.26}$$

Therefore, the centred hat matrix M_{k+1} is symmetric in S_{k+1} and M_k (by (2.25) and (2.26)). In the case $q = 2$ which arises naturally in the modelling of spatial variation in two dimensions, we find

$$\begin{aligned} M_2 &= I - (I - S_2)(I - M_1 S_2)^{-1}(I - M_1) \\ &= I - (I - S_2)(I - S_1 S_2)^{-1}(I - S_1) \end{aligned} \tag{2.27}$$

and thus (2.10) generalises the result given in Hastie and Tibshirani (1990). They give (2.27) as the hat-matrix for the model with two smoothers and no regressor variables.

Remark 2 Here, we summarise the expressions for the estimates of f_{k+1} and $\hat{f}_1 + \ldots + \hat{f}_k$.

$$\begin{aligned} \hat{f}_{k+1} &= (I - S_{k+1}M_k)^{-1}S_{k+1}(I - M_k)y \text{ by (2.14)} \\ &= S_{k+1}(I - M_k S_{k+1})^{-1}(I - M_k)y \text{ by (2.21)} \\ &= (I - (I - S_{k+1}M_k)^{-1}(I - S_{k+1}))y, \end{aligned} \tag{2.28}$$

and

$$\begin{aligned} \hat{f}_1 + \ldots + \hat{f}_k &= (I - M_k S_{k+1})^{-1}M_k(I - S_{k+1})y \\ &= M_k(I - S_{k+1}M_k)^{-1}(I - S_{k+1})y \text{ by (2.18)} \\ &= (I - (I - M_k S_{k+1})^{-1}(I - M_k))y \text{ by (2.22).} \end{aligned}$$

Remark 3 As a last remark, we give two further properties satisfied by the centred hat matrix M_k.

$$M_k 1 = 0 \tag{2.29}$$

$$1' M_k = 0' \tag{2.30}$$

Proof We prove (2.29) by induction,

1. For $k = 1$, $M_1 1 = S_1 1 = 0$ by (2.10) and (2.3).

2. Assume the result is true for $k = 1, \ldots, m$.

3. For $k = m + 1$, by (2.26)

$$\begin{aligned} M_{m+1}1 &= 1 - (I - M_m)(I - S_{m+1}M_m)^{-1}(I - S_{m+1})1 \text{ by (2.26)} \\ &= 1 - (I - M_m)(I - S_{m+1}M_m)^{-1}1 \text{ by (2.3)}, \end{aligned}$$

but $(I - S_{m+1}M_m)^{-1}1 = 1$ since $(I - S_{m+1}M_m)1 = 1$ by the induction hypothesis. Thus, $M_{m+1}1 = 0$ by (2.3).

We omit the proof of (2.30) which is obtained, in a similar way, by induction.

2.2.2 Symmetry of the hat matrix

We have shown that the centred hat matrix M_{k+1} is symmetric in S_{k+1} and M_k. We prove by induction that M_{k+1} **is symmetric if all smoothers are symmetric.**

1. for $i = 1$, $M_1 = S_1$, and is therefore symmetric.

2. Assume M_i is symmetric for $i = 1, \ldots, k$.

3. From (2.25),

$$\begin{aligned} M'_{k+1} &= I - (I - M'_k)(I - S'_{k+1}M'_k)^{-1}(I - S'_{k+1}) \\ &= I - (I - M_k)(I - S_{k+1}M_k)^{-1}(I - S_{k+1}) \\ &= M_{k+1} \text{ by (2.26)}. \end{aligned}$$

Hastie and Tibshirani (1990) made this remark only in the two-smoother case. This property will allow us to reduce the computational cost of calculating standard errors in the semi-parametric case.

2.2.3 Estimation of additive models when all smoothers but one are centred

From the semi-parametric model perspective, it is of interest to explore the case when all but one of the smoothers are centred. We write the additive model (2.11) as

$$y = f_1 + \ldots + f^*_{k+1} + \epsilon, \tag{2.31}$$

where $E(y) = f^*_{k+1}$ and we write $f^*_{k+1} = \alpha + f_{k+1}$. We will assume that the smoother matrix, S^*_{k+1}, associated with f_{k+1} is symmetric, therefore, by (2.1) and (2.2),

$$S^*_{k+1} = S_{k+1} + 11'/n. \tag{2.32}$$

We have in mind the case when S^*_{k+1} is the projection smoother $X(X'X)^{-1}X'$.

Lemma 2.2 *The centred hat matrix, M_k, of an additive model $y = \alpha + f_1 + \ldots + f_k + \epsilon$ with k smooth terms and the non-centred smooth matrix S^*_{k+1} satisfy*

$$M_k S^*_{k+1} = M_k S_{k+1} \tag{2.33}$$

$$11'/n(I - M_k S^*_{k+1})^{-1} = 11'/n \tag{2.34}$$

Proof From the definition of a symmetric centred smooth matrix given in (2.32) we obtain

$$\begin{aligned} M_k S_{k+1} &= M_k(S^*_{k+1} - 11'/n) \\ &= M_k S^*_{k+1} - M_k 11'/n \\ &= M_k S^*_{k+1} \text{ by (2.29).} \end{aligned}$$

(2.34) is immediate, since $(I - M_k S^*_{k+1})11'/n = 11'/n$ from (2.2) and (2.29). ∎

Theorem 2.2 *The estimated smooth terms for model (2.31) are*

$$\hat{f}_1 + \hat{f}_2 + \ldots + \hat{f}_k = [I - (I - M_k S^*_{k+1})^{-1}(I - M_k)]y \tag{2.35}$$

$$\hat{f}^*_{k+1} = S^*_{k+1}(I - M_k S^*_{k+1})^{-1}(I - M_k)y \tag{2.36}$$

$$\hat{y} = [I - (I - S^*_{k+1})(I - M_k S^*_{k+1})^{-1}(I - M_k)]y \tag{2.37}$$

Proof The proof of (2.35) is immediate from (2.22) and (2.33). Further,

$$\begin{aligned} \hat{f}^*_{k+1} &= \hat{\alpha} + \hat{f}_{k+1} \\ &= (11'/n)y + S_{k+1}(I - M_k S_{k+1})^{-1}(I - M_k)y \text{ from (2.21)} \\ &= (11'/n)y + S_{k+1}(I - M_k S^*_{k+1})^{-1}(I - M_k)y \text{ from (2.33)} \\ &= (11'/n)y + (S^*_{k+1} - 11'/n)(I - M_k S^*_{k+1})^{-1}(I - M_k)y \text{ by (2.32)} \\ &= (11'/n)y + S^*_{k+1}(I - M_k S^*_{k+1})^{-1}(I - M_k)y - (11'/n)(I - M_k)y \text{ by (2.34)} \\ &= S^*_{k+1}(I - M_k S^*_{k+1})^{-1}(I - M_k)y \text{ by (2.30).} \end{aligned}$$

Finally, $\hat{y} = \hat{f}_1 + \ldots + \hat{f}_k + \hat{f}^*_{k+1}$, and thus (2.35) and (2.36) yield (2.37) ∎

2.2.4 Existence and uniqueness of the solutions

The conditions for the convergence and independence from the starting values of the solutions of the backfitting algorithm are equivalent to the conditions for the consistency and non-degeneracy of the solutions of the estimating equations of the additive model (2.5). Equation (2.24) proves that if

$$\|S_{k+1}M_k\| < 1, \tag{2.38}$$

the estimating equations are consistent and have a unique solution. Let

$$\|S\| = \|S\|_2 = \max_{1\leq i\leq n} \sqrt{\mu_i},$$

where μ_i are the eigenvalues of $S'S$ (we choose the spectral norm for convenience). If the matrix S is symmetric with positive eigenvalues, the spectral norm coincides with the spectral radius, $\|S\| = \rho(S) = \max_{1\leq i\leq n} |\lambda_i|$, where λ_i are the eigenvalues of S.

Theorem 2.3 *If all the smoothers in the additive model (2.5) are symmetric with eigenvalues in* $[0,1)$, *the normal equations have a unique solution.*

Proof It will be enough to show that $\|M_k\| < 1$. We prove this result by induction.

1. For $i = 1$, $M_1 = S_1$. Since $\rho(S_1) < 1$, $\|M_1\| < 1$.

2. We assume the result is true for $i = 1, \ldots, k$.

3. From equation (2.25) we need to prove that

$$\|I - (I - S_{k+1})(I - M_kS_{k+1})^{-1}(I - M_k)\| < 1,$$

or equivalently,

$$0 < \|(I - S_{k+1})(I - M_kS_{k+1})^{-1}(I - M_k)\| \leq 1.$$

As we showed in section 2.2.2, this matrix is symmetric. Therefore, the norm is equal to the spectral radius. We have that

$$0 < \rho((I - S_{k+1})(I - M_kS_{k+1})^{-1}(I - M_k)) \leq \rho(I - S_{k+1})\rho((I - M_kS_{k+1})^{-1})\rho(I - M_k). \tag{2.39}$$

The first inequality in (2.39) is immediate, since the three matrices are non-singular. Let $\lambda_1 \geq \lambda_2, \ldots, \geq \lambda_n$; $\mu_1 \geq \mu_2, \ldots, \geq \mu_n$ and $\gamma_1 \geq \gamma_2, \ldots, \geq \gamma_n$ be the eigenvalues of $M_k S_{k+1}$, M_k and S_{k+1}. Then, using the properties of the eigenvalues of a symmetric matrix (see for example Schott, 1997), the following relations are satisfied,

$$\begin{aligned} \lambda_1 &\leq \gamma_1 \\ \rho(I - M_k) &= 1 - \mu_n \\ \rho((I - M_k S_{k+1})^{-1}) &= 1/(1-\lambda_1) \leq 1/(1-\gamma_1) \leq 1/(1-\gamma_n). \end{aligned}$$

Hence (2.39) may be rewritten as:

$$0 < \|(I - S_{k+1})(I - M_k S_{k+1})^{-1}(I - M_k)\| \leq (1-\gamma_n)\frac{1}{1-\gamma_n}(1-\mu_n) < 1,$$

and so, $\|M_{k+1}\| < 1$, which ensures the existence of a unique solution to the normal equations. ∎

Some linear smoothers (such as smoothing splines) have eigenvalues in $[0, 1]$, and the eigenvalue equal to 1 has multiplicity 2. The eigenvectors are $\mathbf{1}$, a column vector of ones and $\mathbf{z}$, where z is the covariate against which we are smoothing. Centring the smoother matrix (as we did in the previous section) ensures that $\mathbf{1}$ is not an eigenvector, and so it is necessary to impose some restrictions on the covariates. Now $\|S_{k+1} M_k\| = 1$ if a vector z is reproduced by S_{k+1} and M_k at the same time. This will not happen, unless there is a linear dependence between the covariates. This effect is called **concurvity**.

In conclusion, the solution of the normal equations exists and is unique if the smoothers are symmetric and have eigenvalues in $[0, 1)$. This is equivalent to the case of symmetric, centred smoothers with eigenvalues in $[0, 1]$ and with an empty concurvity space (to ensure that $\|S_{k+1} M_k\| \neq 1$). This coincides with the conditions given by Buja et al. (1989) and extends the conditions for the two-smoother case in which the smoothers could have eigenvalues in $(-1, 1]$. In the case of q smoothers it is necessary to restrict the interval to $[0, 1]$ to ensure that M_k does not have eigenvalues larger than 1. When all smoothers but one are centred, the sufficient condition is $\|M_k S^*_{k+1}\| < 1$. If S^*_{k+1} is the only non-centred matrix, that condition still holds even in the case that

$\|S^*_{k+1}\| = 1$.

Smoothers like locally-weighted running lines (loess) are not symmetric and have eigenvalues larger that 1. However, empirical evidence shows that in many cases the necessary inverses exist giving a unique solution to (2.6) (Hastie and Tibshirani, 1990).

2.3 Semiparametric additive models as a special case of additive models

Semiparametric additive models (**SAMs**) may be seen as a special case of additive models in which one or more terms are linear. SAMs are also an extension of the classic semiparametric model in which all terms but one are linear. We write the model as

$$y = X\beta + f_1 + \ldots + f_q + \epsilon \tag{2.40}$$

where y is the vector of observed values, X is the design matrix (where 1, the vector of 1's, is the column space of X), β is the vector of treatment effects, or more generally of regression coefficients, f_i is a smooth term which is a function of a covariate z_i, $i = 1, \ldots, q$, and ϵ is a vector of independent errors with common variance σ^2. We will refer to ~~model~~ (2.~~40~~) as a semi-parametric model with q smooth terms. We consider (2.40) as an additive model with $q + 1$ terms

$$y = f_0 + f_1 + \ldots + f_q + \epsilon,$$

where $f_0 = X\beta$ and $S_0 = \tilde{X} = X(X'X)^{-1}X'$, the hat matrix of an ordinary linear regression model. Substituting S^*_{k+1} by $\tilde{X}$ in (2.36), (2.35) and (2.37), we obtain expressions for $\hat{\beta}$, $\hat{f}_1 + \ldots + \hat{f}_q$ and $\hat{y}$. First, for $\hat{\beta}$ we have from (2.36)

$$\begin{aligned} X\hat{\beta} &= \tilde{X}(I - M_q\tilde{X})^{-1}(I - M_q)y \\ &= (I - \tilde{X}M_q)^{-1}\tilde{X}(I - M_q)y \quad \text{by (2.19)}, \end{aligned}$$

and so

$$X'(I - \tilde{X}M_q)X\hat{\beta} = X'(I - M_q)y$$

$$X'(I - M_q)X\hat{\beta} = X'(I - M_q)y,$$

which yields

$$\hat{\beta} = (X'(I - M_q)X)^{-1}X'(I - M_q)y. \quad (2.41)$$

Now, for $\hat{f}_1 + \ldots + \hat{f}_q$ we have from (2.35)

$$\hat{f}_1 + \ldots + \hat{f}_q = [I - (I - M_q\tilde{X})^{-1}(I - M_q)]y, \quad (2.42)$$

with

$$\begin{aligned}(I - M_q\tilde{X})^{-1} &= (I - M_qX(X'X)^{-1}X')^{-1} \\ &= I + M_qX(X'(I - M_q)X)^{-1}X' \text{ by (2.19)} \\ &\quad \text{with } D = I,\ B = X',\ A = X'X \text{ and } C = M_qX.\end{aligned}$$

Hence in (2.42) we obtain

$$\hat{f}_1 + \ldots + \hat{f}_q = [M_q - M_qX(X'(I - M_q)X)^{-1}X'(I - M_q)]y = M_q(y - X\hat{\beta}). \quad (2.43)$$

Further, the fitted values for model (2.40) are

$$\begin{aligned}\hat{y} &= X\hat{\beta} + \hat{f}_1 + \ldots + \hat{f}_q \\ &= X\hat{\beta} + M_q(y - X\hat{\beta}) \\ &= M_qy + (I - M_q)X\hat{\beta}. \quad (2.44)\end{aligned}$$

Thus we see that $\hat{y}$ is a weighted average of the observations y and the fitted values from the regression of y on X. The centred matrix M_q has a double role as a weight matrix, first as the weight matrix in (2.44), and second as the weight matrix in the regression of y on X. The hat matrix for model (2.40) is by (2.41) and (2.44)

$$H = M_q + (I - M_q)X(X'(I - M_q)X)^{-1}X'(I - M_q). \quad (2.45)$$

Equation (2.41) is an extension of a well-known result for a semi-parametric model with a single smooth term (see Green et al., 1985; Heckman, 1986)

$$\hat{\beta} = (X'(I - S)X)^{-1}X'(I - S)y$$

provided that $X'(I-S)X$ is invertible. We ensure that the inverse exists by centring the smoother. Green et al. (1985) give an alternative approach to centring; they assume that 1 is in the column space of S but impose the constraint $1'X\beta = 0$. Formally, we can obtain (2.41) by applying the general estimating equations approach of Green et al. (1985). The solution for β to the estimating equations

$$\begin{aligned} f &= M_q(y - X\beta) \\ \beta &= (X'X)^{-1}X'(y - X\beta) \end{aligned} \tag{2.46}$$

is given by (2.41). These are generalised least squares normal equations with a non-diagonal weight matrix $(I - M_q)$. Once M_q is calculated, β is estimated in one step after M_q is computed. This approach resembles the modified back-fitting algorithm proposed by Buja et al. (1989) in which all linear terms in the model and the projection part of the smoothers are combined into a single projection. However the projection step will have to be iterated together with the other steps.

2.3.1 Extension of Speckman's method

Speckman (1988) introduced an alternative method to the penalised least squares approach of Green et al. (1985) for the semi-parametric model (2.40) for the case $q = 1$. We use the same type of argument as in the previous section. Let M_q be the centred smoother matrix of an additive model with q smooth terms, and define the partial residual vectors after adjusting for z_i, $i = 1, \ldots, q$, as $\tilde{y} = (I - M_q)y$ and $\tilde{X} = (I - M_q)X$. We estimate β by applying ordinary least squares to the regression set up

$$(I - M_q)y = (I - M_q)X\beta + \epsilon$$

and $f_1 + \ldots + f_q$ by

$$\hat{f}_1 + \ldots + \hat{f}_q = M_q(y - X\hat{\beta}), \tag{2.47}$$

which yields

$$\begin{aligned} \hat{\beta} &= (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y} \\ \hat{f}_1 + \ldots + \hat{f}_q &= M_q(I - X(\tilde{X}'\tilde{X})^{-1}\tilde{X}'(I - M_q))y. \end{aligned} \tag{2.48}$$

Expression (2.48) corresponds to equation (3.3a) and expression (2.47) to (3.3b) in Speckman (1988).

If M_q is a projection (as in the case of bin smoothers, least squares lines, polynomial regression and regression splines), this method gives the same estimates as the additive model approach.

Speckman (1988) and Severini and Wong (1992) showed that, in the single smoother case, the estimate of β given in (2.48) has smaller bias and is asymptotically more efficient than (2.41). The efficiency of this method in the semiparametric additive case is still to be investigated. Even in the case of a single smoother, Speckman's method is computationally more demanding than the additive model approach. The substantial amount of computation involved in the q-smoother case (since it will be necessary to calculate $M_q'M_q$) and the loss of interpretability make the additive approach more attractive.

2.3.2 Estimation of residual variance in SAMs

Most of the theoretical work in semi-parametric models has concentrated on estimating β. However, an estimate of σ^2 is needed if we want to make inferences on $\hat{\beta}$. Based on the analogy with linear regression, we define $\hat{\sigma}^2$ (Buckley et al., 1988) as

$$\hat{\sigma}^2 = \frac{y'(I-H)'(I-H)y}{Trace((I-H)'(I-H))} = \frac{y'(I-H)'(I-H)y}{n - Trace(2H - H'H)}, \tag{2.49}$$

where H is the hat matrix given in (2.45) and $n - Trace(H' + H - H'H)$ are the *equivalent degrees of freedom* for the error term. This estimate of σ^2 is unbiased if $M_q(f_1 + \ldots + f_q) = f_1 + \ldots + f_q$ (which extends the condition for unbiasedness given in Green et al., 1985). However, the expression for $\hat{\sigma}^2$ given in (2.49) is computationally very demanding and we prefer to use an estimate based on the one given by Wahba

(1983)

$$\hat{\sigma}^2 = \frac{y'(I-H)'(I-H)y}{n - Trace(H)}. \tag{2.50}$$

We will assume that the conditions given in section 2.2.4 are satisfied, that is, all the smoothers are symmetric with eigenvalues in $[0,1]$, then H is symmetric (by the result given in section 2.2.2). Let λ_i $i = 1, \ldots, n$ be the eigenvalues of H. Standard results for symmetric matrices yield the following relationships:

$$\begin{aligned} Trace(H) &= \sum_{i=1}^{n} \lambda_i \\ Trace(HH) &= \sum_{i=1}^{n} \lambda_i^2 \\ Trace(2H - HH) &= \sum_{i=1}^{n} 2\lambda_i - \lambda_i^2. \end{aligned}$$

We have proved in Theorem 2.3 that if the eigenvalues of S_i are in $[0,1]$, the eigenvalues of the hat matrix are also in that interval, therefore,

$$Trace(HH) \leq Trace(H) \leq Trace(2H - HH). \tag{2.51}$$

This suggests that (2.50) may underestimate the residual variance. A small simulation study suggests that (2.51) may also hold for loess. However, the number of calculations necessary to compute $Trace(H)$ is still very large. As an alternative, we use the sum of the degrees of freedom of the individual smoothers, since it gives an upper bound on the degrees of freedom of the fitted model (Buja et al., 1989) (this approximation is exact if the covariates are linearly independent). Thus, we take

$$\hat{\sigma}^2 = \frac{y'(I-H)'(I-H)y}{n - p - d}, \tag{2.52}$$

where $d = \sum_{i=1}^{q-1} d_i$, and d_i are the degrees of freedom of each S_i and p is the rank of X.

2.4 Standard errors in SAMs

The main parameter of interest in a semi-parametric additive model is often β, and the average mean squared error in estimated treatment differences is reported as a measure of accuracy of the experiment. In previous sections we derived an explicit expression for the value of $\hat{\beta}$ which will allow us to calculate the standard error of $\hat{\beta}$.

2.4.1 The calculation of the standard errors

From equation (2.41) we can obtain the following expression for the variance of $\hat{\beta}$

$$Var(\hat{\beta}) = AA'\sigma^2, \tag{2.53}$$

where

$$A = (X'(I - M_q)X)^{-1}X'(I - M_q). \tag{2.54}$$

Hence the estimation of $Var(\hat{\beta})$ requires the value of M_q and an estimate of σ^2. We use the estimate of σ^2 given in (2.52) and we find M_q with a well-known device (Hastie and Tibshirani, 1987): we fit the n models $e_j = \alpha + f_1 + \ldots + f_q + \epsilon$, where e_j are the columns of the $n \times n$ identity matrix I_n. The fitted values for these n models give M_q^* and M_q follows from (2.7). The variance of $\hat{\beta}$ depends on M_q through M_qX and $X'M_q$, therefore, when all the smoothers are symmetric, M_q is symmetric and a knowledge of M_qX is sufficient to calculate $Var(\hat{\beta})$ from (2.53). To find M_qX we modify the device used to calculate M_q as follows: we replace the columns of I_n with the columns of X to obtain M_q^*X and from there, $M_qX = (I - 11'/n)M_q^*X$. Thus when M_q is symmetric we need to fit $p < n$ models (p being the rank of X).

Cubic smoothing splines are symmetric smoothers so in this case we can calculate $Var(\hat{\beta})$ exactly if we know M_qX. Then, the matrix A given in (2.53) can be calculated by

$$A = (X'(X - B))^{-1}(X - B)', \; B = M_qX. \tag{2.55}$$

2.4.2 Two approximations for the calculation of the standard errors

The expression for the variance of $\hat{\beta}$ based on (2.55) only holds exactly for symmetric smoothers. We propose the use of (2.55) even when M_q is not symmetric, as for example, in the case of Cleveland's (Cleveland, 1979) locally-weighted running line (loess). Our approximation is based on the assumption that $M_q'X \sim M_qX$ and this allows $Var(\hat{\beta})$ to be calculated by fitting only p models. We also consider a simple approximation which is used, for example, in the Genstat package (Genstat-5, 1995):

suppose that Z is the matrix whose columns are the covariates in the smooth part of the model. Define X^* to be X augmented with Z; that is, we add the linear part of the smoother to the regression part of the model. This definition of X^* is motivated by the use of the modified back-fitting algorithm in which the smoother is divided into a linear and a non-linear (or shrinking) part, and the linear part is fitted together with the other linear terms in the model. The approximation consists of assuming that the non-linear part of M_q is orthogonal to X, and thus we can compute $Var(\hat{\beta})$ with

$$Var(\hat{\beta}) = (X'^* X^*)^{-1}\sigma^2. \tag{2.56}$$

The advantage of this approximation is that the calculations required are the ones for an ordinary linear model. However in the examples presented in the next section we show that when the amount of smoothing needed is considerable, this approximation may underestimate considerably the variance of $\hat{\beta}$.

2.4.3 Some examples

We examine the two approximations in four data sets which reflect our interest in the application of semi-parametric additive models in agriculture, but otherwise are quite general. Suppose we model the yields of a crop in an agricultural trial with model (2.40) where X is the design matrix (with rank p) corresponding to variety effects, and the smooth terms correspond to fertility effects which vary slowly across the field. The value of q (the number of smooth terms) will often be 1 (a single two-dimensional smoother) or 2 (two one dimensional smoothers), although our final example has $q = 2$ for loess and $q = 3$ in the case of smoothing splines.

We assess the approximations by considering both the mean percentage error and the maximum percentage error in the standard error of the difference (SED) in the pairwise comparisons between variety effects. Suppose M_q is the centred smoother matrix corresponding to the smooth part of model (2.40), and take the degrees of freedom associated with M_q to be $Trace(M_q) = d$. For degrees of freedom d, we define $s_i(d)$ to be the SED computed from (2.53) and (2.54) (where i varies over

all pairwise comparisons), $a_{i1}(d)$ is our approximate SED calculated from (2.53) and (2.55) and $a_{i2}(d)$ is the approximate SED calculated from (2.56). Thus,

$$u_j(d) = \frac{1}{m}\sum_i \frac{s_i(d) - a_{ij}(d)}{s_i(d)} \qquad j = 1, 2$$

$$v_j(d) = \max_i \frac{s_i(d) - a_{ij}(d)}{s_i(d)} \qquad j = 1, 2$$

where $i = 1, \ldots, m$.

When smoothing splines are used in model (2.40), the calculation of SED with (2.54) and (2.55) is exact, so we only need to consider the approximation given by (2.56). We plot $u_j(d)$ and $v_j(d)$ against the degrees of freedom d used to estimate the smooth part of the model for both smoothers, loess (Figure 2.1) and smoothing splines (Figure 2.2). One general point is that these plots only depend on X and M_q, i.e., only on the design of the trial and not on the observed values y. In Table 2.1 we compare the approximations through $u_j(d^*)$ and $v_j(d^*)$ where d^* is the preferred degrees of freedom, chosen by minimising the generalised cross-validation (GCV) score originally introduced by Craven and Wahba (1979). Of course, d^* depends on the yields y, so the appropriate position on the plot for assessing the adequacy of either approximation is not known until y has been observed.

Example 1: Barley trial, Rothamsted (1979)

Jenkyn et al. (1979) present data from a study on mildew control with $n = 36$ and four treatments representing different spray frequencies. The plots were arranged in a single row, in nine blocks of four plots. Green et al. (1985) used least squares smoothing to analyse these data and Hastie and Tibshirani (1987) use the additive model approach and reported the covariance matrix for the treatment estimates. We do the smoothing with 1-dimensional loess and a single cubic smoothing spline. In Figure 2.1 we see that for loess, our approximation to the SED is very good for all degrees of freedom and, in particular for $d = d^*$, giving a maximum percentage error of 0.2%. The mean and maximum percentage error with approximation (2.56) increases as the degrees of freedom increase and gave a mean percentage error of 11.6% with

			Loess					Splines		
Example	n	p	d^*	$u_1(d^*)$	$u_2(d^*)$	$v_1(d^*)$	$v_2(d^*)$	d^*	$u_2(d^*)$	$v_2(d^*)$
1	36	4	14.9	-0.02	11.6	0.2	13.2	15	9.6	10.1
2	150	25	38.5	-0.06	7.2	1.0	11.2	13	2.9	7.1
3	360	25	50.0	-0.03	3.8	0.4	6.4	13(5)	0.5	2.1
4	320	6	41.0	-0.07	4.5	0.1	6.8	20	3.5	4.8

Table 2.1: *Sample size, n, number of regressors, p, preferred smooth degrees of freedom, d^*, (figure in () is degrees of freedom for rows only), mean percentage error $u_j(d^*)$, and maximum percentage error, $v_j(d^*)$, in standard error of difference for variety effects.*

loess at $d = d^*$, while with splines it gave an error of 9.6% at $d = d^*$. We also note that approximation (2.56) gives similar errors for both loess and splines with similar degrees of freedom.

Example 2: Spring wheat trial, Slate Hall Farm (1976)

Gilmour et al. (1995) give yields of 25 varieties of wheat, laid out in six replicates in a balanced lattice square design and so $n = 150$ and $p = 25$ in this example. A plot of residuals from the analysis of variance against row and column position suggested the presence of a spatial trend. Further residual plots, using conditional plots (Cleveland, 1994) suggested that a two-dimensional smoother might be more appropriate. With loess, we use a two-dimensional smooth surface, while we use two 1-dimensional smoothers for splines (since two-dimensional splines are not readily available). The preferred degrees of freedom with loess was $d^* = 38.5$ (equivalent to a neighbourhood of 12 points). Figure 2.1 and Table 2.1 indicate that our approximation performed very well. For loess we found a mean SED of 67 both with (2.54) and (2.55) and 62 with (2.56) (Gilmour et al., 1995 reported a mean SED of 62 $grams/m^2$). The plot in Figure 2.2 is not well-defined, because with two 1-dimensional splines, we can obtain any degrees of freedom d for the smooth part of the model in many different ways. For a given d value we have plotted the mean/maximum percentage error

for the model that corresponds to minimising the generalised cross-validation score conditional on d. For splines, $d^* = 13$ and the mean SED was 94 with (2.55) and 91 with (2.56). These large values reflect the poor fit of an additive model with two 1-dimensional smoothers; however, the poor fit does not invalidate the comparison of the approximations.

Example 3: Barley trial, Scottish Crop Research Institute (1995)

Hackett and Newton (1995) give details of the layout of a barley trial with 20 varieties, three levels of fungicide and 2 levels of nitrogen and so $p = 25$. The trial was a row and column design (by which we mean a design in which rows and columns are orthogonal) with 20 rows and 18 columns. The chosen degrees of freedom with loess were $d^* = 50$ giving a mean percentage error of -0.03 (with maximum value of 0.4) for our approximation and 6.4 with approximation (2.56). With splines, approximation (2.56) seems to perform well with a maximum percentage error of 2.1%, however we need to take into account that the number of degrees of freedom in this case is 13, much smaller than $d^* = 50$ used with loess. In general, in a row and column layout as we described above, the maximum number of degrees of freedom used for the nonparametric part of an additive model with a smoother for row position and another for column position will be $d_1 + d_2$, where d_1 is the number of rows and d_2 the number of columns. The reason why this occurs is because in an additive model with these two covariates, the hat matrix has the form,

$$M_2 = S_1 + S_2, \tag{2.57}$$

a consequence of the fact that the covariates are exactly balanced, that is, each of the values for rows occur with each of the values for columns, and so

$$S_1 S_2 = 0 \quad \text{(Hastie and Tibshirani, 1990).}$$

This, together with (2.27) yields (2.57). In this particular design, treatments are also orthogonal to columns and so $X'S_2 = 0$. Thus, the matrix A given in (2.54) becomes

$$A = (X'(I - S_1)X)^{-1}X'(I - S_1) \tag{2.58}$$

which does not depend on the model for columns (although $\hat{\sigma}^2$ does). In Table 2.1 the degrees of freedom for rows is reported in brackets and in Figure 2.2 we plot the percentage error in the SED against the row degrees of freedom.

In this example, approximation (2.56) seems to work better with splines than with loess. This is because (2.56) will work better when the smooth term is close to linear and, in general, the smaller the degrees of freedom associated with the smooth part of the model, the better approximation (2.56) will perform. In the case of loess, the degrees of freedom are large, which means the model for rows and columns is far from linear and the approximation works relatively poorly. In the case of splines, the degrees of freedom are smaller and the approximation is also helped because the treatments are orthogonal to columns, i.e., A only depends on the smoother for rows. However, the success of the approximation in this example is not to be confused with a satisfactory fit of the model, since in this example, a model with two one-dimensional splines fails to describe the data satisfactorily.

Example 4: Winter barley trial, SCRI (1996)

The data are from part of a large barley trial carried out at the Scottish Crop Research Institute in 1996. We will analyse this trial in more detail in chapter 5. The part of the trial analysed consisted of 16 beds grouped into 4 blocks of four beds. The trial was laid out in 16 columns and 20 rows and so $n = 320$. Each block contained two main plots of two beds which received a fungicide treatment (absence/presence) allocated randomly within blocks. Each bed was divided into 20 sub-plots and 5 cultivars were allocated at random to alternate sub-plots. Check plots of a sixth variety were grown between the cultivar plots. The trial was also scored for mildew and the area under disease (mildew) progress curve, AUDPC, was used as a covariate; see Carver and Griffiths (1981).

For loess, a model with the six main effects for varieties and two smooth terms (one two-dimensional smoother for rows and columns and the one-dimensional smoother with AUDPC as a covariate) described the data well. We found $d^* = 41$ and our

approximation worked well, giving a mean percentage error of -0.07% and a maximum of 0.05%. The approximation (2.56) was less successful with $u_2(d^*) = 4.5\%$ and $v_2(d^*) = 6.8\%$; as in the previous examples, we do not expect this approximation to work well since the model is very far from linear ($d^* = 41$). In this example, the total degrees of freedom of the smooth part of the model does not uniquely determine the model. In Figure 2.1 we have plotted the mean/maximum percentage difference in the SED for the model that gives the minimum cross-validation score conditional on d; some smoothing of the resulting plot was also done.

With splines, we use three one-dimensional smoothers. Again, Figure 2.2 was produced in the same way as for the loess smoother. The number of degrees of freedom is 20 and approximation (2.56) performs better in this case than for the loess smoother. However, the model with two loess smoothers gave a much better fit to the data.

Remark In all cases, approximation (2.53) and (2.55) performed better than (2.56), requires little extra effort and is exact in the case of splines. This approximation has been used on a number of other examples and the results agree with the ones shown here, suggesting that the approximation may be used with confidence.

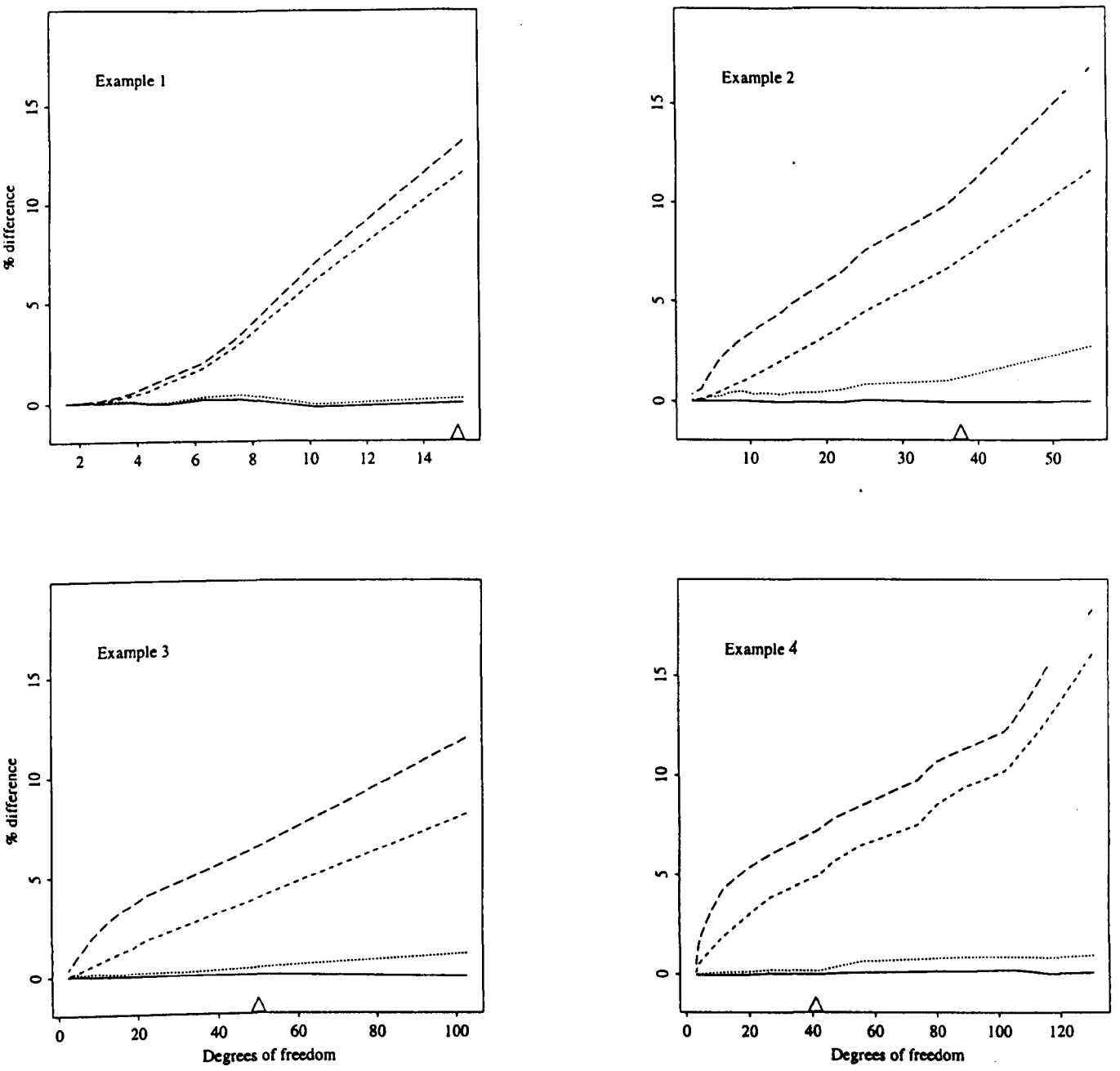

Figure 2.1: *Plot of percentage error in the standard error of difference in pairwise comparisons versus degrees of freedom. The mean and maximum percentage for approximation* (2.55) $(u_1(\delta^*), v_1(\delta^*))$ *are represented by* —— *and* ....., *while the mean and maximum percentage for approximation* (2.56) $(u_2(\delta^*), v_2(\delta^*))$ *are represented by* - - - - - *and* — — — —. *The preferred degrees of freedom is indicated by* Δ.

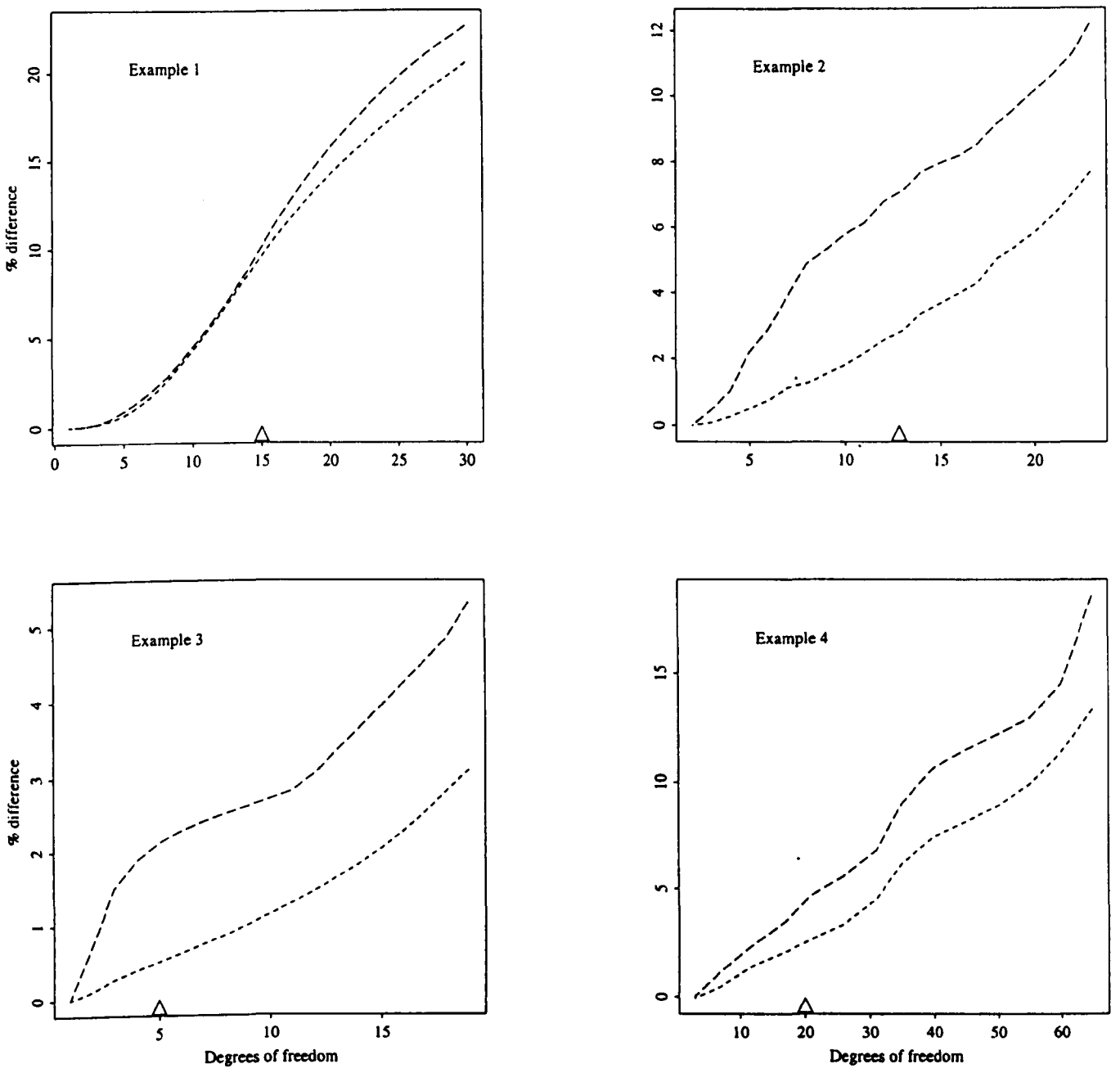

Figure 2.2: *Plot of percentage error in the standard error of difference in pairwise comparisons versus degrees of freedom for the smooth part of the model. The mean and maximum percentage for approximation* (2.56) $(u(d^*), v(d^*))$ *are represented by* - - - - - *and* ———. *The preferred degrees of freedom is indicated by* Δ.

Chapter 3

Exact adjustment of the profile likelihood for a class of normal regression models

3.1 Introduction

There are several approaches to the problem of inference in the presence of nuisance parameters. In the case of regression models, when the parameters of interest are *variance parameters*, methods such as the modified (or restricted) likelihood of Patterson and Thompson (1971) which estimates variance parameters by means of the likelihood for a set of error contrasts, or the marginal likelihood of Kalbfleisch and Sprott (1970) are widely used. However, in more general cases, e.g, regression models in which the parameters of interest are in the mean and in the variance simultaneously, these methods fail. One simple and general approach is to form the *profile likelihood* by maximising the nuisance parameters for fixed values of the parameters of interest and then replace them in the likelihood function. In general, the assumption that we can eliminate the nuisance parameters by the use of the profile likelihood leads to bias, inconsistency and over-optimistic variance estimates, especially when the number of nuisance parameters is large.

A number of authors have suggested different modifications to the profile likelihood in order to improve the asymptotic behavior of the estimates and their standard errors, e.g., the modified profile likelihood of Barndorff-Nielsen (1986) and the conditional profile likelihood of Cox and Reid (1987) (though these versions of the profile likelihood are based on an assumption of orthogonality between the nuisance parameters and the parameters of interest). Cox and Reid (1993) extended their method to cases in which parameters are not orthogonal. The approach we take here is the method introduced by McCullagh and Tibshirani (1990) which attempts to adjust the profile log-likelihood so that the mean of the score function is zero and the variance is equal to minus the expectation of its derivative matrix. McCullagh and Tibshirani (1990) gave two approximations to the calculation of these adjustments. In Section 3.3, we give exact adjustments for the profile likelihood for a class of regression models and show that, when the parameters of interest are only present in the variance, the adjusted profile likelihood is equivalent to the restricted maximum likelihood of Patterson and Thompson (1971). In Section 3.4, we prove that the approximate bias adjustments of Cox and Reid (1993) and McCullagh and Tibshirani (1990) are exact for our class of regression models. We illustrate our method with three examples.

3.2 Notation and definitions

We will use the notation of McCullagh and Tibshirani (1990). We assume that the distribution of the random variable Y depends on a parameter $\theta' = (\psi', \lambda')$ where $\psi' = (\psi_1, \ldots, \psi_r)$ is the parameter of interest and $\lambda' = (\lambda_1, \ldots \lambda_p)$ is the nuisance parameter. Let $y' = (y_1, \ldots, y_n)$ be a vector of observations on Y, and denote the log-likelihood by $\ell(\theta) = \ell(\psi, \lambda)$. The maximum likelihood estimate of θ is $\hat{\theta}' = (\hat{\psi}', \hat{\lambda}')$. The profile log-likelihood $\ell_{\mathrm{p}}(\psi)$ is obtained by replacing the nuisance parameters λ with their maximum likelihood estimates $\hat{\lambda}_\psi$, for fixed values of the parameter of interest ψ, that is,

$$\ell_{\mathrm{p}}(\psi) = \ell(\psi, \hat{\lambda}_\psi).$$

The profile log-likelihood score function, $U(\psi)$, is given by

$$U(\psi) = \frac{\partial}{\partial \psi} \ell_{\mathrm{p}}(\psi).$$

Following McCullagh and Tibshirani (1990), we define

$$m(\psi) = \mathrm{E}\left(\frac{\partial \ell_{\mathrm{p}}}{\partial \psi}\right), \tag{3.1}$$

$$W_1(\psi) = \mathrm{Var}\left(\frac{\partial \ell_{\mathrm{p}}}{\partial \psi}\right) \tag{3.2}$$

$$W_2(\psi) = -E\left(\frac{\partial^2 \ell_{\mathrm{p}}}{\partial \psi^2} - \frac{\partial m}{\partial \psi}\right). \tag{3.3}$$

McCullagh and Tibshirani adjust $U(\psi)$ so that it has mean zero and its variance is minus the expected derivative matrix. They take expectations in (3.1), (3.2) and (3.3) at $(\psi, \hat{\lambda}_\psi)$; for simplicity, we take expectations at (ψ, λ) and give all results in this way. Replacing λ by $\hat{\lambda}_\psi$ recovers McCullagh and Tibshirani's approach.

Definitions

The adjusted profile score is defined as

$$\tilde{U}(\psi) = W_2(\psi)' W_1(\psi)^{-1}(U(\psi) - m(\psi)). \tag{3.4}$$

The score function is said to be **unbiased** if

$$\mathrm{E}_{\psi,\lambda}\left(\tilde{U}(\psi)\right) = 0, \tag{3.5}$$

and we will say it is **information unbiased** when

$$\mathrm{Var}_{\psi,\lambda}\left(\tilde{U}(\psi)\right) = -\mathrm{E}_{\psi,\lambda}\left(\frac{\partial \tilde{U}(\psi)}{\partial \psi}\right). \tag{3.6}$$

Lemma 3.1 *The adjusted profile score $\tilde{U}(\psi)$ given in (3.4) is unbiased and information unbiased.*

Proof

$$\mathrm{E}_{\psi,\lambda}\left(\tilde{U}(\psi)\right) = W_2(\psi)' W_1(\psi)^{-1}[\mathrm{E}_{\psi,\lambda}(U(\psi)) - m(\psi)]$$

$$
\begin{aligned}
&= W_2(\psi)'W_1(\psi)^{-1}(m(\psi) - m(\psi)) \\
&= 0,
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Var}_{\psi,\lambda}\left(\tilde{U}(\psi)\right) &= W_2(\psi)'W_1(\psi)^{-1}\mathrm{Var}_{\psi,\lambda}\left(U(\psi) - m(\psi)\right)W_1(\psi)^{-1}W_2(\psi) \\
&= W_2(\psi)'W_1(\psi)^{-1}W_1(\psi)W_1(\psi)^{-1}W_2(\psi) \\
&= W_2(\psi)'W_1(\psi)^{-1}W_2(\psi),
\end{aligned}
$$

$$
\begin{aligned}
-\mathrm{E}_{\psi,\lambda}\left(\frac{\partial\tilde{U}(\psi)}{\partial\psi}\right) &= -\left(\frac{\partial W_2(\psi)'W_1(\psi)^{-1}}{\partial\psi}\right)\mathrm{E}_{\psi,\lambda}\left(U(\psi) - m(\psi)\right) \\
&\quad -W_2(\psi)'W_1(\psi)^{-1}\mathrm{E}_{\psi,\lambda}\left(\frac{\partial^2\ell_{\mathrm{p}}}{\partial\psi^2} - \frac{\partial m}{\partial\psi}\right) \\
&= 0 + W_2(\psi)'W_1(\psi)^{-1}W_2(\psi) \\
&= \mathrm{Var}_{\psi,\lambda}\left(\tilde{U}(\psi)\right).
\end{aligned}
$$

■

Of course, implicit in the proof of this lemma is the fact that $m(\psi)$, $W_1(\psi)$ and $W_2(\psi)$ do not depend on λ and therefore on y. We will give examples in which all these conditions are satisfied and indicate a possible solution for the cases when $W_1(\psi)$ and $W_2(\psi)$ depend on λ.

In the case of a scalar parameter, the adjusted profile log-likelihood is defined as

$$\ell_{\mathrm{ap}} = \int^{\psi} \tilde{U}(\psi)\, dt.$$

The calculation of the adjusted profile score requires the computation of $m(\psi)$, $W_1(\psi)$ and $W_2(\psi)$. In some cases, it is difficult to give analytical expressions for these quantities, so McCullagh and Tibshirani proposed two methods of approximation: Monte Carlo simulation and a first-order approximation for the adjustments given in terms of the cumulants of the derivatives of the unadjusted score function.

We restrict our attention to a class of normal regression models which cover a wide range of examples and for which we can calculate the exact adjustments. The models

are defined by

$$y \sim \mathcal{N}(X(\psi)\lambda, \Sigma(\psi)). \tag{3.7}$$

The class of models defined in (3.7) contains many examples such as, linear mixed models, the many normal means problem, the non-linear regression model discussed by Macaskill (1993) and the competition models of Draper and Guttman (1980) and Kempton (1982) among others (see Durban and Currie, 1998 for further examples). For model (3.7), we define

$$G(\psi) = (X'(\psi)\Sigma(\psi)^{-1}X(\psi))^{-1} \tag{3.8}$$

$$H(\psi) = \Sigma(\psi)^{-1}X(\psi)G(\psi)X'(\psi)\Sigma(\psi)^{-1} \tag{3.9}$$

$$Q(\psi) = \Sigma(\psi)^{-1} - H(\psi). \tag{3.10}$$

Usually we will suppress the dependence of $X(\psi)$, $\Sigma(\psi)$, $G(\psi)$, etc., on ψ, and write them simply as X, Σ, G, etc.

The computations involved in the calculation of $\tilde{U}(\psi)$ can be very laborious, however we will show in the next section that it is possible to simplify them. For example, the calculation of $W_2(\psi)$ appears to require

$$\frac{\partial^2 Q}{\partial\psi_i\partial\psi_j}.$$

Macaskill (1993) gives an expression for this matrix for a particular case of model (3.7) in the appendix of his paper. We will show how the calculation of this matrix can be avoided, giving a simplified alternative expression.

3.3 Calculation of the adjusted score

The full log-likelihood for model (3.7) is

$$\ell(\psi,\lambda) = -\frac{1}{2}\log|\Sigma| - \frac{1}{2}(y - X\lambda)'\Sigma^{-1}(y - X\lambda). \tag{3.11}$$

For ψ fixed, the solution of $\partial\ell/\partial\lambda = 0$ gives the result

$$\hat{\lambda}_\psi = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y. \tag{3.12}$$

Substituting this expression into (3.11), we obtain the profile log-likelihood

$$\ell_{\rm p}(\psi) = -\frac{1}{2}\log|\Sigma| - \frac{1}{2}y'Qy, \tag{3.13}$$

where Q is given in (3.10).

The derivation of $m(\psi)$, $W_1(\psi)$ and $W_2(\psi)$ involves the first two derivatives of $\ell_{\rm p}(\psi)$ with respect to ψ, i.e., we need the first two derivatives of Q. In the following lemma we give several identities satisfied by Q and its derivatives.

Lemma 3.2 *With G and Q defined in (3.8) and (3.10) respectively we have for scalar parameters ψ, ψ_i and ψ_j*

$$Q\Sigma Q = Q; \tag{3.14}$$

$$QX = 0 \text{ and } \frac{\partial Q}{\partial \psi}X = -Q\frac{\partial X}{\partial \psi}; \tag{3.15}$$

$$\frac{\partial Q}{\partial \psi} = -Q\frac{\partial \Sigma}{\partial \psi}Q - Q\frac{\partial X}{\partial \psi}GX'\Sigma^{-1} - \Sigma^{-1}XG\frac{\partial X'}{\partial \psi}Q; \tag{3.16}$$

$$tr\left(\frac{\partial Q}{\partial \psi}\Sigma\right) = -tr\left(\frac{\partial \Sigma}{\partial \psi}Q\right); \tag{3.17}$$

$$tr\left(\frac{\partial^2 Q}{\partial \psi_i \partial \psi_j}\Sigma\right) = -tr\left(\frac{\partial^2 \Sigma}{\partial \psi_i \partial \psi_j}Q + \frac{\partial \Sigma}{\partial \psi_i}\frac{\partial Q}{\partial \psi_j} + \frac{\partial Q}{\partial \psi_i}\frac{\partial \Sigma}{\partial \psi_j}\right); \tag{3.18}$$

$$X'\frac{\partial^2 Q}{\partial \psi_i \partial \psi_j}X = \frac{\partial X'}{\partial \psi_i}Q\frac{\partial X}{\partial \psi_j} + \frac{\partial X'}{\partial \psi_j}Q\frac{\partial X}{\partial \psi_i}. \tag{3.19}$$

Proof (3.14) follows immediately from the definition of Q given in (3.10). It is also immediate that $QX = 0$ and taking derivatives in this expression we have

$$\frac{\partial Q}{\partial \psi}X + Q\frac{\partial X}{\partial \psi} = 0$$

and so (3.15) follows. Differentiating (3.14) with respect to ψ we obtain

$$\frac{\partial Q}{\partial \psi} = \frac{\partial Q}{\partial \psi}\Sigma Q + Q\frac{\partial \Sigma}{\partial \psi}Q + \left(\frac{\partial Q}{\partial \psi}\Sigma Q\right)', \tag{3.20}$$

and from the definition of Q

$$\Sigma Q = I - XGX'\Sigma^{-1}. \tag{3.21}$$

The substitution of (3.21) into (3.20) yields

$$\frac{\partial Q}{\partial \psi} = 2\frac{\partial Q}{\partial \psi} - \frac{\partial Q}{\partial \psi} XGX'\Sigma^{-1} + Q\frac{\partial \Sigma}{\partial \psi} Q - \left(\frac{\partial Q}{\partial \psi} XGX'\Sigma^{-1}\right)',$$

and using (3.15) we obtain (3.16).

To prove (3.17) we show that $\mathrm{tr}(Q\Sigma) = n - p$ and differentiate this expression with respect to ψ. We have

$$\mathrm{tr}\,(Q\Sigma) = \mathrm{tr}\left(I - \Sigma^{-1}XGX'\right) = n - \mathrm{tr}\left(GX'\Sigma^{-1}X\right) = n - p.$$

Then,

$$\mathrm{tr}\left(\frac{\partial Q}{\partial \psi}\Sigma + \frac{\partial \Sigma}{\partial \psi} Q\right) = 0$$

which gives (3.17). Equation (3.18) is obtained by differentiating (3.17).

Finally to prove (3.19) we note that from (3.15)

$$X'\frac{\partial Q}{\partial \psi_j} X = 0,$$

and differentiating with respect to ψ_i we get

$$\frac{\partial X'}{\partial \psi_i}\frac{\partial Q}{\partial \psi_j} X + X'\frac{\partial^2 Q}{\partial \psi_i \partial \psi_j} X + X'\frac{\partial Q}{\partial \psi_j}\frac{\partial X}{\partial \psi_i} = 0,$$

and using (3.15) again we obtain

$$-\frac{\partial X'}{\partial \psi_i} Q\frac{\partial X}{\partial \psi_j} + X'\frac{\partial^2 Q}{\partial \psi_i \partial \psi_j} X - \frac{\partial X'}{\partial \psi_j} Q\frac{\partial X}{\partial \psi_i} = 0,$$

which yields (3.19). ∎

Now, we are able to calculate the bias and information adjustment suggested by (McCullagh and Tibshirani, 1990). First, we give three matrix results (see Schott, 1997):

For $z \sim \mathcal{N}(\mu, V)$ and C and D matrices of constants,

$$\frac{\partial \log|C|}{\partial x} = \mathrm{tr}\left(C^{-1}\frac{\partial C}{\partial x}\right) \tag{3.22}$$

$$\mathrm{E}(z'Cz) = \mathrm{tr}(CV) + \mu'C\mu \tag{3.23}$$

$$\mathrm{Cov}(z'Cz, z'Dz) = 2\mathrm{tr}(CVDV) + 4\mu'CVD\mu. \tag{3.24}$$

We calculate the bias adjustment in the following lemma.

Lemma 3.3 *If* $m(\psi) = m = (m_1, \ldots, m_r)'$ *then*

$$m_i = -\frac{1}{2} tr\left(\frac{\partial \Sigma}{\partial \psi_i} H\right). \tag{3.25}$$

Proof For convenience and without loss of generality we will assume that ψ is scalar. Differentiating (3.13) and using (3.22), we find

$$\frac{\partial \ell_p}{\partial \psi} = -\frac{1}{2}\mathrm{tr}\left(\frac{\partial \Sigma}{\partial \psi}\Sigma^{-1}\right) - \frac{1}{2} y' \frac{\partial Q}{\partial \psi} y. \tag{3.26}$$

Now take expectations and use (3.23). We find

$$\begin{aligned} \mathrm{E}\left(\frac{\partial \ell_p}{\partial \psi}\right) &= -\frac{1}{2}\mathrm{tr}\left(\frac{\partial \Sigma}{\partial \psi}\Sigma^{-1}\right) - \frac{1}{2}\mathrm{tr}\left(\frac{\partial Q}{\partial \psi}\Sigma\right) - \frac{1}{2}\lambda' X' \frac{\partial Q}{\partial \psi} X\lambda \\ &= -\frac{1}{2}\mathrm{tr}\left(\frac{\partial \Sigma}{\partial \psi}\Sigma^{-1}\right) + \frac{1}{2}\mathrm{tr}\left(\frac{\partial \Sigma}{\partial \psi} Q\right) \quad \text{by (3.17) and (3.15).} \end{aligned}$$

The result follows from the definition of $Q = \Sigma^{-1} - H$. ∎

Remark The bias adjustment does not depend on the nuisance parameter λ and therefore does not depend on the vector of observed values y. There is no bias correction when the covariance matrix Σ does not depend on ψ. The form of the correction $m(\psi)$ is the same whether X depends on ψ or not. This implies that there will be a bias correction only if the parameter of interest is present in the covariance matrix. When the parameter of interest it is in the covariance matrix only, we can use a REML type of adjustment for model (3.7). We will show that when X does not depend on ψ the REML adjustment and $m(\psi)$ coincide. However, when ψ is present in X, it is not possible to find a set of error contrasts $z = Ty$ where T does not depend on ψ. Hence, it is necessary to use an alternative approach, such as McCullagh and Tibshirani's.

In the next lemma we give the expression for $W_1(\psi)$.

Lemma 3.4 *The (i,j) th element of W_1 is*

$$W_1(i,j) = \frac{1}{2}tr\left(Q\frac{\partial\Sigma}{\partial\psi_i}Q\frac{\partial\Sigma}{\partial\psi_j} + 2G\frac{\partial X'}{\partial\psi_i}Q\frac{\partial X}{\partial\psi_j}\right) + \lambda'\frac{\partial X'}{\partial\psi_i}Q\frac{\partial X}{\partial\psi_j}\lambda. \tag{3.27}$$

Proof From the definition of $W_1(\psi)$ given in (3.2) and the expression for $\partial\ell_p/\partial\psi$ given in (3.26) we have

$$\begin{aligned} W_1(i,j) &= \mathrm{Cov}\left(-\frac{1}{2}\mathrm{tr}\left(\frac{\partial\Sigma}{\partial\psi_i}\Sigma^{-1}\right) - \frac{1}{2}y'\frac{\partial Q}{\partial\psi_i}y, -\frac{1}{2}\mathrm{tr}\left(\frac{\partial\Sigma}{\partial\psi_j}\Sigma^{-1}\right) - \frac{1}{2}y'\frac{\partial Q}{\partial\psi_j}y\right) \\ &= \frac{1}{4}\mathrm{Cov}\left(y'\frac{\partial Q}{\partial\psi_i}y, y'\frac{\partial Q}{\partial\psi_j}y\right) \text{ since } \mathrm{tr}\left(\frac{\partial\Sigma}{\partial\psi_j}\Sigma^{-1}\right) \text{ is a constant.} \end{aligned}$$

We use (3.24) to obtain

$$W_1(i,j) = \frac{1}{2}\mathrm{tr}\left(\frac{\partial Q}{\partial\psi_i}\Sigma\frac{\partial Q}{\partial\psi_j}\Sigma\right) + \lambda'X'\frac{\partial Q}{\partial\psi_i}\Sigma\frac{\partial Q}{\partial\psi_j}X\lambda. \tag{3.28}$$

Now from (3.16) we have

$$\frac{\partial Q}{\partial\psi}\Sigma = -Q\frac{\partial\Sigma}{\partial\psi}Q\Sigma - Q\frac{\partial X}{\partial\psi}GX' - \Sigma^{-1}XG\frac{\partial X'}{\partial\psi}Q\Sigma,$$

and so

$$\begin{aligned} \mathrm{tr}\left(\frac{\partial Q}{\partial\psi_i}\Sigma\frac{\partial Q}{\partial\psi_j}\Sigma\right) &= \mathrm{tr}\left(Q\frac{\partial\Sigma}{\partial\psi_i}Q\Sigma\left(Q\frac{\partial\Sigma}{\partial\psi_j}Q\Sigma + \underbrace{Q\frac{\partial X}{\partial\psi_j}GX'}_{0} + \underbrace{\Sigma^{-1}XG\frac{\partial X'}{\partial\psi_j}Q\Sigma}_{0}\right)\right) \\ &\quad + \mathrm{tr}\left(Q\frac{\partial X}{\partial\psi_i}GX'\left(\underbrace{Q\frac{\partial\Sigma}{\partial\psi_j}Q\Sigma}_{0} + \underbrace{Q\frac{\partial X}{\partial\psi_j}GX'}_{0} + \Sigma^{-1}XG\frac{\partial X'}{\partial\psi_j}Q\Sigma\right)\right) \\ &\quad + \mathrm{tr}\left(\Sigma^{-1}XG\frac{\partial X'}{\partial\psi_i}Q\Sigma\left(\underbrace{Q\frac{\partial\Sigma}{\partial\psi_j}Q\Sigma}_{0} + Q\frac{\partial X}{\partial\psi_j}GX' + \underbrace{\Sigma^{-1}XG\frac{\partial X'}{\partial\psi_j}Q\Sigma}_{0}\right)\right) \\ &= \mathrm{tr}\left(Q\frac{\partial\Sigma}{\partial\psi_i}Q\frac{\partial\Sigma}{\partial\psi_j} + Q\frac{\partial X}{\partial\psi_i}G\frac{\partial X'}{\partial\psi_j}Q\Sigma + G\frac{\partial X'}{\partial\psi_i}Q\frac{\partial X}{\partial\psi_j}\right) \\ &= \mathrm{tr}\left(Q\frac{\partial\Sigma}{\partial\psi_i}Q\frac{\partial\Sigma}{\partial\psi_j} + Q\frac{\partial X}{\partial\psi_i}G\frac{\partial X'}{\partial\psi_j} + G\frac{\partial X'}{\partial\psi_i}Q\frac{\partial X}{\partial\psi_j}\right) \\ &= \mathrm{tr}\left(Q\frac{\partial\Sigma}{\partial\psi_i}Q\frac{\partial\Sigma}{\partial\psi_j} + 2Q\frac{\partial X}{\partial\psi_i}G\frac{\partial X'}{\partial\psi_j}\right). \end{aligned} \tag{3.29}$$

We also have,

$$\lambda'X'\frac{\partial Q}{\partial\psi_i}\Sigma\frac{\partial Q}{\partial\psi_j}X\lambda = \lambda'\frac{\partial X'}{\partial\psi_i}Q\frac{\partial X}{\partial\psi_j}\lambda \quad \text{by (3.15).} \tag{3.30}$$

Substituting (3.29) and (3.30) into (3.28) we obtain the expression for $W_1(\psi)$ given in (3.27). ■

Remark When X does not depend on ψ, the expression for $W_1(\psi)$ simplifies and does not depend on λ (this is the case of the linear mixed model).

Lemma 3.5 *The (i,j) th element of W_2 is*

$$W_2(i,j) = \frac{1}{2}tr\left(Q\frac{\partial\Sigma}{\partial\psi_i}Q\frac{\partial\Sigma}{\partial\psi_j} + 2\Sigma^{-1}XG\frac{\partial X'}{\partial\psi_i}Q\frac{\partial\Sigma}{\partial\psi_j}\right) + \lambda'\frac{\partial X'}{\partial\psi_i}Q\frac{\partial X}{\partial\psi_j}\lambda. \qquad (3.31)$$

Proof W_2 is defined in (3.3) as

$$\begin{aligned}
W_2(i,j) &= -\mathrm{E}\left(\frac{\partial}{\partial\psi_j}\left(\frac{\partial\ell_p}{\partial\psi_i} - m_i(\psi)\right)\right) \\
&= \frac{1}{2}\mathrm{E}\left(\frac{\partial}{\partial\psi_j}\left(\mathrm{tr}\left(\frac{\partial\Sigma}{\partial\psi_i}\Sigma^{-1}\right) + y'\frac{\partial Q}{\partial\psi_i}y - \mathrm{tr}\left(\frac{\partial\Sigma}{\partial\psi_i}H\right)\right)\right) \quad \text{by (3.26) and (3.25)} \\
&= \frac{1}{2}\mathrm{E}\left(\frac{\partial}{\partial\psi_j}\left(\mathrm{tr}\left(\frac{\partial\Sigma}{\partial\psi_i}Q\right) + y'\frac{\partial Q}{\partial\psi_i}y\right)\right) \quad \text{by (3.10)} \\
&= \frac{1}{2}\mathrm{tr}\left(\frac{\partial^2\Sigma}{\partial\psi_i\partial\psi_j}Q + \frac{\partial\Sigma}{\partial\psi_i}\frac{\partial Q}{\partial\psi_j}\right) + \frac{1}{2}\mathrm{E}\left(y'\frac{\partial^2 Q}{\partial\psi_i\partial\psi_j}y\right) \\
&= \frac{1}{2}\mathrm{tr}\left(\frac{\partial^2\Sigma}{\partial\psi_i\partial\psi_j}Q + \frac{\partial\Sigma}{\partial\psi_i}\frac{\partial Q}{\partial\psi_j} + \frac{\partial^2 Q}{\partial\psi_i\partial\psi_j}\Sigma\right) + \frac{1}{2}\lambda'X'\frac{\partial^2 Q}{\partial\psi_i\partial\psi_j}X\lambda \qquad (3.32) \\
&\quad \text{by (3.23).}
\end{aligned}$$

Both terms in (3.32) involve the second derivative of Q which is a complicated calculation as shown by Macaskill (1993) in the appendix to his paper. We avoid the calculation of this matrix using some of the equalities given in Lemma 3.2. The term in λ becomes

$$\begin{aligned}
\frac{1}{2}\lambda'X'\frac{\partial^2 Q}{\partial\psi_i\partial\psi_j}X\lambda &= \frac{1}{2}\lambda'\left(\frac{\partial X'}{\partial\psi_i}Q\frac{\partial X}{\partial\psi_j} + \frac{\partial X'}{\partial\psi_j}Q\frac{\partial X}{\partial\psi_i}\right)\lambda \quad \text{by (3.19)} \\
&= \lambda'\frac{\partial X'}{\partial\psi_i}Q\frac{\partial X}{\partial\psi_j}\lambda. \qquad (3.33)
\end{aligned}$$

The first term in (3.32) reduces to

$$\frac{1}{2}\mathrm{tr}\left(\frac{\partial^2\Sigma}{\partial\psi_i\partial\psi_j}Q + \frac{\partial\Sigma}{\partial\psi_i}\frac{\partial Q}{\partial\psi_j} + \frac{\partial^2 Q}{\partial\psi_i\partial\psi_j}\Sigma\right) = -\frac{1}{2}\mathrm{tr}\left(\frac{\partial Q}{\partial\psi_i}\frac{\partial\Sigma}{\partial\psi_j}\right) \quad \text{by (3.18)}$$

and

$$-\frac{1}{2}\mathrm{tr}\left(\frac{\partial\Sigma}{\partial\psi_j}\frac{\partial Q}{\partial\psi_i}\right) = \frac{1}{2}\mathrm{tr}\left(\frac{\partial\Sigma}{\partial\psi_j}\left(Q\frac{\partial\Sigma}{\partial\psi_i}Q + Q\frac{\partial X}{\partial\psi_i}GX'\Sigma^{-1} + \Sigma^{-1}XG\frac{\partial X'}{\partial\psi_i}Q\right)\right) \quad \text{by (3.16)}$$

$$= \frac{1}{2}\mathrm{tr}\left(Q\frac{\partial\Sigma}{\partial\psi_i}Q\frac{\partial\Sigma}{\partial\psi_j} + 2\Sigma^{-1}XG\frac{\partial X'}{\partial\psi_i}Q\frac{\partial\Sigma}{\partial\psi_j}\right). \tag{3.34}$$

The result follows substituting (3.33) and (3.34) in (3.32). ∎

Remark First, as we pointed out before, the identities given in Lemma 3.2 avoid calculating the second derivative of Q simplifying the calculation of W_2. Second, W_2 is not symmetric in general. However there are two important cases where it is symmetric. If X or Σ do not depend on ψ, W_2 is symmetric. More importantly, when X does not depend on ψ, W_2 does not depend on λ and $W_1 = W_2$,-i.e., no variance adjustment is necessary. When W_1 and W_2 depend on λ and we replace λ by λ_ψ we will indicate this by writing the adjustments as $W_1^*(\psi)$ and $W_2^*(\psi)$.

Now, we summarise the previous lemmas in the following theorem.

Theorem 3.1 *Suppose* $y \sim \mathcal{N}(X(\psi)\lambda,\ \Sigma(\psi))$ *where* $\psi = (\psi_1, \ldots, \psi_r)'$ *is the parameter of interest. Let* $U = (U_1, \ldots, U_r)'$ *be the score function where* $U_i = \partial\ell_p/\partial\psi_i$ *is given in (3.26).*

(i) If the adjusted score function is defined by

$$\tilde{U} = U - m \tag{3.35}$$

where m *is defined in (3.25) then* $\tilde{U}$ *is unbiased. Moreover, if* Σ *does not depend on* ψ *then* $m = 0$ *and the profile log-likelihood,* ℓ_p*, is unbiased.*

(ii) If the adjusted score function is defined by

$$\tilde{U} = W_2'W_1^{-1}(U - m) \tag{3.36}$$

where m*,* W_1 *and* W_2 *are defined in (3.25), (3.27) and (3.31) respectively then* $\tilde{U}$ *is unbiased and information unbiased. Moreover, if* X *does not depend on* ψ *then* $W_1 = W_2$ *and the adjusted profile log-likelihood,* ℓ_{ap}*, is given by*

$$\ell_{ap} = -\frac{1}{2}\log|\Sigma| - \frac{1}{2}\log|X'\Sigma^{-1}X| - \frac{1}{2}y'Qy. \tag{3.37}$$

Proof From Lemma 3.1 and (3.35) $\tilde{U}$ is unbiased. The bias adjustment

$$m_i = -\frac{1}{2}\left(\frac{\partial \Sigma}{\partial \psi_i} H\right)$$

only depends on the derivative of the variance Σ with respect to ψ, therefore if Σ does not depend on ψ, the unadjusted score U is unbiased.

The first part of (ii) follows from Lemma 3.1. The final part follows from (3.22). Since X does not depend on ψ,

$$\begin{aligned}
\frac{\partial}{\partial \psi_i} \log |X'\Sigma^{-1}X| &= -\mathrm{tr}\left(X'\Sigma^{-1}\frac{\partial \Sigma}{\partial \psi_i}\Sigma^{-1}X\left(X'\Sigma^{-1}X\right)^{-1}\right) \\
&= -\mathrm{tr}\left(\frac{\partial \Sigma}{\partial \psi_i}\Sigma^{-1}X\left(X'\Sigma^{-1}X\right)^{-1}X'\Sigma^{-1}\right) \\
&= -\mathrm{tr}\left(\frac{\partial \Sigma}{\partial \psi_i}H\right) \quad \text{by (3.9)}
\end{aligned}$$

and the second term in (3.37) gives rise to the bias adjustment m. ∎

3.4 Approximations to the adjusted profile likelihood

In this section we consider two other approaches to the inference of a vector of parameters in the presence of nuisance parameters. Cox and Reid (1987) describe a method of adjustment for the profile likelihood in which the nuisance parameters are required to be orthogonal to the parameters of interest and they present a method for the construction of orthogonal parameters. This method involves a set of partial differential equations which it is not possible to solve in some cases and therefore this method cannot be used in all situations. Cox and Reid (1993) proposed an approximation to the adjusted profile likelihood which did not require that the nuisance parameters be orthogonal to the parameters of interest. Here, we consider this approximation together with the first-order approximations to the calculation of the bias adjustment $m(\psi)$, and the scale adjustment, $W(\psi)$ given by McCullagh and Tibshirani (1990).

3.4.1 Cox and Reid approximation

Cox and Reid (1993) give an approximate adjustment for the profile likelihood in terms of the original parameterisation. The adjusted profile likelihood is given in their equation (5) as

$$\ell_{\mathrm{a}}(\psi) = \ell(\psi, \hat{\lambda}_\psi) - \frac{1}{2}\log|j_{\lambda\lambda}| + (\psi - \psi_0)m^r_{,r} \tag{3.38}$$

where the third term in this equation is the trace of the matrix $\partial m^r/\partial\lambda^u$, the derivative being evaluated at $(\hat{\psi}, \hat{\lambda})$ and ψ_0 replaced by $\hat{\psi}$.

We have used the notation of Cox and Reid, except that their nuisance parameter ϕ has been replaced by our parameter λ. We now calculate the quantities in (3.38). First we differentiate the full log-likelihood (3.11) for model (3.7) to obtain

$$\begin{aligned}
j_{\lambda\lambda} &= \frac{\partial^2\ell}{\partial\lambda^2} = X'\Sigma^{-1}X \\
m_r &= i^{rs}i_{0s} \text{ as in their equation (3)} \\
i_{rs} &= -\mathrm{E}\left(\frac{\partial^2\ell}{\partial\lambda_r\partial\lambda_s}\right) = \left(X'\Sigma^{-1}X\right)_{rs} \\
i_{0s} &= -\mathrm{E}\left(\frac{\partial^2\ell}{\partial\psi\partial\lambda_s}\right) = \left(X'\Sigma^{-1}\frac{\partial X}{\partial\psi}\lambda\right)_s .
\end{aligned}$$

We find $m^r_{,r}$ as the trace of $\partial m^r/\partial\lambda^u$; we obtain

$$m^r_{,r} = \mathrm{tr}\left(\left(X'\Sigma^{-1}X\right)^{-1}X'\Sigma^{-1}\frac{\partial X}{\partial\psi}\right). \tag{3.39}$$

Equation (5) in Cox and Reid (1993) yields their equation (7)

$$\ell_N = \ell_{\mathrm{p}} - \frac{1}{2}\log|X'\Sigma^{-1}X| + (\psi - \hat{\psi})\,\mathrm{tr}\left(\left(X'\Sigma^{-1}X\right)^{-1}X'\Sigma^{-1}\frac{\partial X}{\partial\psi}\right) \tag{3.40}$$

the last term being evaluated at $(\hat{\psi}, \hat{\lambda})$. Differentiating with respect to ψ and assuming $\psi - \hat{\psi}$ is small we find

$$\frac{\partial\ell_N}{\partial\psi} = \frac{\partial\ell_{\mathrm{p}}}{\partial\psi} - \frac{1}{2}\frac{\partial\log|X'\Sigma^{-1}X|}{\partial\psi} + \mathrm{tr}\left(\left(X'\Sigma^{-1}X\right)^{-1}X'\Sigma^{-1}\frac{\partial X}{\partial\psi}\right)$$

and

$$\frac{\partial\log|X'\Sigma^{-1}X|}{\partial\psi} = 2\mathrm{tr}\left(\left(X'\Sigma^{-1}X\right)^{-1}X'\Sigma^{-1}\frac{\partial X}{\partial\psi}\right) - \mathrm{tr}\left(\frac{\partial\Sigma}{\partial\psi}H\right) \quad \text{from (3.22).}$$

Then,

$$\frac{\partial \ell_N}{\partial \psi} = \frac{\partial \ell_{\mathrm{p}}}{\partial \psi} + \frac{1}{2}\left(\frac{\partial \Sigma}{\partial \psi} H\right).$$

Thus Cox and Reid (1993) produces the same adjustment to the profile log-likelihood score as found in Lemma 3.3.

Remark If the parameter of interest is not in the span of X, λ and ψ are orthogonal and $m^r_{,r} = 0$. Thus, equation (3.38) coincides with the modified profile log-likelihood given in Cox and Reid (1987). When $\partial X/\partial \psi \neq 0$, the last term in equation (3.40) is a linear approximation of the correction term which allows us to calculate the value of the adjusted profile log-likelihood in the neighbourhood of the true parameter.

3.4.2 McCullagh and Tibshirani approximation

McCullagh and Tibshirani (1990) proposed the adjustment for the profile likelihood score that we are using throughout this chapter. In general, the bias adjustment, $m(\psi)$, and the scale adjustment, $W(\psi)$, require formidable calculations, and McCullagh and Tibshirani provide first order approximations to these adjustments which involve derivatives of the full log-likelihood instead of the profile log-likelihood. We show that, for model (3.7), the first-order bias adjustment given by McCullagh and Tibshirani (1990) is exact, but the variance adjustment is not. First, we introduce the index notation used in McCullagh and Tibshirani (1990).

The components of ψ are denoted by ψ^r and the derivatives of the log-likelihood with respect to ψ and λ are denoted by:

$$U_r = \frac{\partial \ell}{\partial \psi^r} \quad U_i = \frac{\partial \ell}{\partial \lambda^i} \quad U_{ij} = \frac{\partial^2 \ell}{\partial \lambda^i \lambda^j} \quad U_{ri} = \frac{\partial^2 \ell}{\partial \psi^r \lambda^i}$$

and the cumulants of these derivatives:

$$\begin{aligned} k_{ri} &= \mathrm{E}(U_{ri}) = -\mathrm{Cov}(U_r, U_i) = -\mathrm{E}(U_r U_i) = -k_{r,i} \\ k_{ij} &= \mathrm{E}(U_{ij}) = -\mathrm{E}(U_i U_j) = -k_{i,j} \\ k^{i,j} &= k^{-1}_{i,j}. \end{aligned}$$

There are two reasons why the first-order approximation is exact in the normal regression case. First, McCullagh and Tibshirani (1990) use the Taylor expansion for the profile log-likelihood up to the quadratic term

$$\ell_{\mathrm{p}} = \ell(\psi, \hat{\lambda}_\psi) = \ell(\psi, \lambda) + \frac{\partial \ell}{\partial \lambda}(\hat{\lambda}_\psi - \lambda) + \frac{1}{2}\frac{\partial^2 \ell}{\partial \lambda^2}(\hat{\lambda}_\psi - \lambda)^2 + \ldots \tag{3.41}$$

but this is exact for the normal regression case, since

$$U_{ijk} = \frac{\partial^3 \ell}{\partial \lambda^i \lambda^j \lambda^k} = 0.$$

Second, they use the approximation for $(\hat{\lambda}_\psi - \lambda)$ given in McCullagh (1987)

$$\hat{\lambda}^i_\psi - \lambda^i = \kappa^{i,j} U_j + \kappa^{i,j}\kappa^{k,l}(U_{jk} - \kappa_{jk})U_l + \frac{1}{2}\kappa^{ijk} U_j U_k + \ldots \tag{3.42}$$

which in our case reduces to

$$\hat{\lambda}^i_\psi - \lambda^i = \kappa^{i,j} U_j \tag{3.43}$$

since

$$k_{jk} = E(U_{jk}) = U_{jk} = \left((X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\right)_{jk} \tag{3.44}$$

and so (3.43) is also exact. McCullagh and Tibshirani substitute (3.42) into the expansion for the derivative of (3.41) with respect to ψ^r. We find this expression for the normal regression case by substituting (3.43) into (3.41); we find

$$\frac{\partial \ell_{\mathrm{p}}}{\partial \psi^r} = U_r + U_{ri}k^{i,j}U_j + \frac{1}{2}U_{rij}k^{i,k}k^{j,l}U_k U_l. \tag{3.45}$$

Taking expectations in this expression we find

$$\mathrm{E}\left(\frac{\partial \ell_{\mathrm{p}}}{\partial \psi^r}\right) = k_{ri,j}k^{i,j} + \frac{1}{2}k_{rij}k^{i,j} \tag{3.46}$$

since $\mathrm{E}(U_r) = 0$ and $k_{ri,j} = \mathrm{Cov}(U_{ri}U_j) = \mathrm{E}(U_{ri}U_j)$. The third-order Bartlett identity

$$\kappa_{rij} = -\kappa_{r,i,j} - \kappa_{r,ij} - \kappa_{ri,j} - \kappa_{rj,i}$$

becomes in this case

$$\kappa_{rij} = -\kappa_{r,i,j} - 2k_{ri,j}. \tag{3.47}$$

Therefore, substituting (3.47) in (3.46) we obtain the exact bias adjustment for the profile likelihood score, what McCullagh and Tibshirani (1990) call the *simplified first-order bias adjustment*

$$m(\psi^r) = -\frac{1}{2}k_{r,i,j}k^{i,j}. \tag{3.48}$$

McCullagh and Tibshirani substitute (3.43) in (3.41), differentiate with respect to ψ, and take expectations. This is precisely the calculation carried out in Lemma 3.3, then, (3.48) is the bias adjustment calculated in Lemma 3.3.

The variance correction is not exact. The first order approximations for W_1 and W_2 given by McCullagh and Tibshirani (1990) are

$$W_1(r,s) = \kappa_{r,s} - \frac{1}{2}\kappa_{r,i,j}\kappa_{s,k,l}\kappa^{i,k}\kappa^{j,l} + O(n^{-1})$$

and

$$W_2(r,s) = W_1(r,s) - \frac{1}{2}\kappa_{r,s,i}\kappa_{j,k,l}\kappa^{i,j}\kappa^{k,l} + O(n^{-1}).$$

The third-order Bartlett identity gives

$$\kappa_{j,k,l} = -\kappa_{jkl} - 3\kappa_{j,kl} = 0$$

since both κ_{jkl} and $\kappa_{j,kl}$ are zero for our class of models. Thus $W_1 = W_2$ up to $O(n^{-1})$ and no scale adjustment is required for our class of models. This suggests that the scale adjustment given in Lemmas 3.4 and 3.5 is likely to be small, as we will confirm in our examples.

3.5 Examples

There are many examples which fit into the class of regression models that we are concerned with. We present only the models that are of most interest to us and have a connection with the rest of the thesis. We have divided the models into three categories, (i) models with known X, (ii) models with independent and normally distributed errors and (iii) a general transformation model.

3.5.1 Models with known X

When ψ is not in the span of X, model (3.7) becomes $Y \sim \mathcal{N}(X\lambda, \Sigma(\psi))$; this is the standard linear mixed model. The usual approach to this type of regression model is the residual log-likelihood of Patterson and Thompson (1971); they estimate

the variance parameters by maximising the likelihood of a set of error contrasts and Harville (1974) showed that using only error contrasts to make inferences on variance components is equivalent to ignoring any prior information on the fixed effects and using all the data. The method gives the modified or restricted log-likelihood for Σ:

$$\ell_{\mathrm{r}} = -\frac{1}{2}\log|\Sigma| - \frac{1}{2}\log|X'\Sigma^{-1}X| - \frac{1}{2}y'Qy. \tag{3.49}$$

We now obtain the McCullagh and Tibshirani adjustments. Lemma 3.3 shows that the bias adjustment is

$$m_i = -\frac{1}{2}\mathrm{tr}\left(\frac{\partial\Sigma}{\partial\psi_i}(\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\right) = \frac{1}{2}\frac{\partial}{\partial\psi_i}\log|X'\Sigma^{-1}X|.$$

Lemmas 3.4 and 3.5 give the scale adjustment in terms of

$$W_1(i,j) = W_2(i,j) = \frac{1}{2}\mathrm{tr}\left(Q\frac{\partial\Sigma}{\partial\psi_i}Q\frac{\partial\Sigma}{\partial\psi_j}\right).$$

Thus, part (ii) of Theorem 4.1 shows that $\ell_{\mathrm{ap}} = \ell_{\mathrm{r}}$ exactly; the residual log-likelihood is thus unbiased and information unbiased.

McCullagh and Tibshirani (1990) prove in their example 8 that the first-order bias-corrected profile log-likelihood is identical to the residual log-likelihood. We have improved this result and show that both likelihoods are identical.

3.5.2 Models with independent and identically distributed errors

We consider the special case of our model where $\Sigma = \sigma^2 I$ and $X(\psi)$ does not depend on σ^2, i.e., model (3.7) becomes $Y \sim \mathcal{N}(X(\psi)\lambda, \sigma^2 I)$.

In this example we take the parameter of interest as $(\psi', \sigma^2)' = (\psi_1, \ldots, \psi_{r-1}, \sigma^2)'$ and applying Lemma 3.3 we find

$$m_1 = m_2 = \ldots = m_{r-1} = 0$$

since there is no bias adjustment when the parameter is only in the span of X; and

$$m_r = -\frac{p}{2\sigma^2}$$

where p is the rank of X. Now,

$$Q = \frac{1}{\sigma^2}(I - P) \text{ where } P = X(X'X)^{-1}X' \quad \text{and } X = X(\psi) \tag{3.50}$$

and so by (3.13) the profile log-likelihood is

$$\ell_{\mathrm{p}} = -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}y'(I - P)y \tag{3.51}$$

and the bias adjusted profile log-likelihood is

$$\ell_{\mathrm{ap}} = -\frac{n-p}{2}\log\sigma^2 - \frac{1}{2\sigma^2}y'(I - P)y. \tag{3.52}$$

It follows that the estimates of $(\psi_1, \ldots, \psi_{r-1})$ based on the adjusted profile log-likelihood are equal to the maximum likelihood estimates based on the full log-likelihood. The estimate of σ^2 is

$$\hat{\sigma}^2 = \frac{y'(I - P)y}{n - p}$$

where P is defined in (3.50). Thus $\hat{\sigma}^2$ is a REML style estimate. The estimates of the ψ_i are the solutions to the system of non-linear equations

$$\frac{1}{2\sigma^2}y'\frac{\partial P}{\partial \psi_i}y = 0 \;\Rightarrow\; y'(I - P)\frac{\partial X}{\partial \psi_i}GX'y = 0,\ i = 1, \ldots, r-1.$$

Lemmas 3.4 and 3.5 together with (3.50) show that

$$W_1(\psi_i, \psi_j) = \mathrm{tr}\left((X'X)^{-1}\frac{\partial X'}{\partial \psi_i}(I - P)\frac{\partial X}{\partial \psi_j}\right) + \frac{1}{\sigma^2}\lambda'\frac{\partial X'}{\partial \psi_i}(I - P)\frac{\partial X}{\partial \psi_j}\lambda \tag{3.53}$$

$$W_1(\psi_i, \sigma^2) = W_1(\sigma^2, \psi_i) = 0 \tag{3.54}$$

$$W_1(\sigma^2, \sigma^2) = \frac{n-p}{2\sigma^4} \tag{3.55}$$

and

$$W_2(\psi_i, \psi_j) = \frac{1}{\sigma^2}\lambda'\frac{\partial X'}{\partial \psi_i}(I - P)\frac{\partial X}{\partial \psi_j}\lambda. \tag{3.56}$$

$$W_2(\psi_i, \sigma^2) = W_2(\sigma^2, \psi_i) = 0 \tag{3.57}$$

$$W_2(\sigma^2, \sigma^2) = \frac{n-p}{2\sigma^4}. \tag{3.58}$$

Macaskill (1993) discussed this model but there he assumes that σ^2 is known and equal to its maximum likelihood estimate. As he remarks, *"the case when σ^2 is unknown yields an awkward expectation of a ratio of quadratic forms which does not*

lead to a simple expression for the correction factor". Having σ^2 as a nuisance parameter implies that the canonical parametrisation is not used and that leads to more complicated calculations. We avoid this problem by considering σ^2 as a parameter of interest. This yields simple calculations and the estimate of σ^2 is adjusted in the manner of a REML estimate.

We discuss an example of this model. Draper and Guttman (1980) suggested a general model for response surface problems in which a treatment is assumed to have a direct effect on the plot to which it is applied and a neighbouring effect on each neighbouring plot. One example of their model may be written

$$Y = G(Z_1\beta + Z_2\gamma) + \epsilon = GZ\lambda + \epsilon \qquad (3.59)$$

where Z_1 and β are the design matrix and parameter vector for the treatment effects, and Z_2 and γ are the design matrix and parameter vector for the block effects. We take $Z = [Z_1 : Z_2]$ and $\lambda' = (\beta', \gamma')$. We take $G = I + \rho S$ where S is the neighbour incidence matrix and ρ is the coefficient of *interference*. We concentrate on the estimation of σ^2 and ρ, and consider λ as a nuisance parameter. Draper and Guttman (1980) analyse data from the mildew control experiment given in Jenkyn et al. (1979). We analysed this data set in the previous chapter and although a semi-parametric model fits the data better, we use this data set here and in the next section for its simplicity. This experiment was set up in a single strip of 36 plots and two border plots to achieve neighbour balance. The maximum likelihood estimates obtained from the adjusted score are $\hat{\rho} = 0.158$ and $\hat{\sigma}^2 = 0.0262$ with standard errors 0.055 and 0.0071 respectively. The value of the scale adjustment factor W (when λ is replaced with $\lambda_{\hat{\psi}}$) is

$$W^*(\hat{\psi}) = \begin{pmatrix} 0.982 & 0 \\ 0 & 1 \end{pmatrix}$$

and again the scale adjustment to the score is very small (of order $O(n^{-1})$). In most cases it is difficult to obtain the adjusted profile log-likelihood for ρ and σ^2 jointly, but noting that ρ and σ^2 are orthogonal, we can take $\sigma^2 = \hat{\sigma}^2 = 0.0262$ and then scale the bias-adjusted score by $w = 0.982$ (Stafford (1996) suggests replacing $w(\psi)$

with $w(\hat{\psi})$, on the basis that they are asymptotically equivalent). An approximate bias and scale adjusted profile log-likelihood for ρ is

$$\ell_{\mathrm{ap}}(\rho) = 0.982\left(\frac{n-p}{2} - \frac{1}{2\hat{\sigma}^2}y'(I-P)y\right). \tag{3.60}$$

Figure 3.1 is a plot of this adjusted profile log-likelihood and the profile log-likelihood in (3.51) with σ^2 replaced by $\hat{\sigma}^2$ where the maximum values of both ℓ_{p} and ℓ_{ap} have been set to be zero.

3.5.3 A general transformation model

The examples in this section are more complicated than in the previous sections, since the parameter(s) of interest are present in the span of X and in the covariance matrix Σ simultaneously. Suppose we observe y and that there exists a transformation

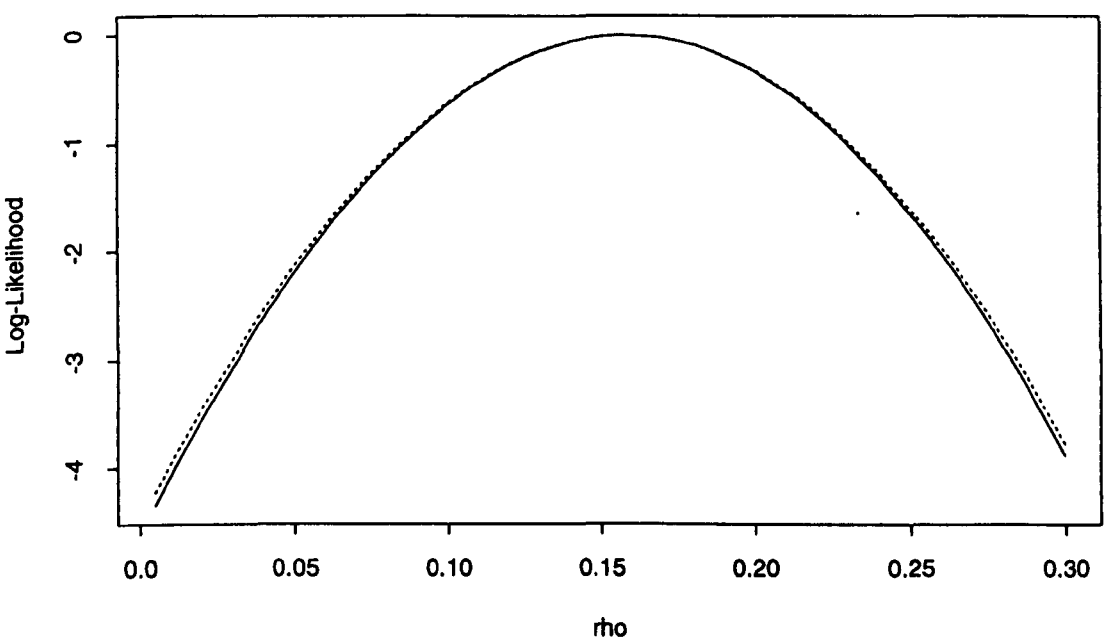

Figure 3.1: *Plot of the profile log-likelihood —– and adjusted profile log-likelihood - - - for the Draper and Guttman example.*

$y \to Dy = z$ such that $z \sim \mathcal{N}(X\lambda, \sigma^2 I)$. Suppose further that $D = D(\rho)$ where ρ is a scalar parameter. Then

$$y \sim \mathcal{N}(D^{-1}X\lambda, \sigma^2 D^{-1}D'^{-1}) \tag{3.61}$$

is an example of our general model with $\psi' = (\rho, \sigma^2)$. The matrices G and Q of (3.8) and (3.10) are

$$G = \sigma^2(X'X)^{-1} \text{ and } Q = \frac{1}{\sigma^2}D'(I - P)D$$

where

$$P = X(X'X)^{-1}X' \tag{3.62}$$

and the profile log-likelihood is

$$\ell_{\mathrm{p}}(\rho, \sigma^2) = -\frac{n}{2}\log\sigma^2 + \log|D| - \frac{1}{2\sigma^2}y'D'(I - P)Dy. \tag{3.63}$$

Direct application of Lemma 3.3 gives

$$m' = \left(\mathrm{tr}(A), -\frac{p}{2\sigma^2}\right) \tag{3.64}$$

where

$$A = P\frac{\partial D}{\partial \rho}D^{-1}. \tag{3.65}$$

The estimate of σ^2 from the adjusted profile likelihood is

$$\hat{\sigma}^2 = \frac{y'D'(I - P)Dy}{n - p}$$

which is a REML style estimate, but based on the transformed observations Dy; in practice, D is replaced by $\hat{D} = D(\hat{\rho})$.

The variance correction matrix $W = W_2'W_1^{-1}$ is found from Lemmas 3.4 and 3.5 to be

$$\begin{aligned}
W_1(\rho, \rho) &= \mathrm{tr}(B^2) + \mathrm{tr}(B'B) + \frac{1}{\sigma^2}\lambda'X'B'BX\lambda \\
W_2(\rho, \rho) &= W_1(\rho, \rho) + \mathrm{tr}(AB) \\
W_1(\rho, \sigma^2) &= W_1(\sigma^2, \rho) = W_2(\rho, \sigma^2) = W_2(\sigma^2, \rho) = -\frac{1}{\sigma^2}\mathrm{tr}(B) \\
W_1(\sigma^2, \sigma^2) &= W_2(\sigma^2, \sigma^2) = \frac{n - p}{2\sigma^4},
\end{aligned}$$

where A is given in (3.65) and

$$B = (I - P)\frac{\partial D}{\partial \rho}D^{-1}.$$

In this case, W_1 and W_2 are symmetric, although this is not true in general as we remarked in section 3.3.

We apply these results to the competition model proposed by Kempton (1982). This model intends to correct for competition effects in yield trials by regression of plot yields on the yields of neighbours, using the mean of the neighbouring plots as a covariate. We use the Jenkyn et al. (1979) data again. Let y_{i-1}, y_i and y_{i+1} be consecutive values in the trial layout; Kempton suggested that

$$y_i = \tau_{[i]} + \beta_{[i]} + \frac{1}{2}\rho(y_{i-1} + y_{i+1}) + \epsilon_i \tag{3.66}$$

where $\tau_{[i]}$ represents the treatment effect for the treatment applied to plot i, and $\beta_{[i]}$ represents the block effect for the block containing plot i. Therefore, the matrix D equals $I - \rho W$, where W has off-diagonal, $(i, i \pm 1)$, elements $1/2$ but is otherwise 0. This matrix is adjusted to allow the yield of border plots to be used for nearest-neighbour adjustment. The expressions involved in the calculation of the bias and variance adjustment given above simplify substantially because $\partial D/\partial \rho = -W$ and $D^{-1}W = WD^{-1}$.

The unadjusted maximum likelihood estimates are $\hat{\sigma}^2 = 0.0203$ and $\hat{\rho} = 0.231$ with a standard error of 0.092. The adjusted values are $\hat{\sigma}^2 = 0.0305$ and $\hat{\rho} = 0.336$ with standard error 0.114. In this case, W^* at the adjusted maximum likelihood estimates for ρ and σ^2 is

$$W^*(\hat{\psi}) = \begin{pmatrix} 1.0349 & 0.00026 \\ 0 & 1 \end{pmatrix}.$$

Since $W^*(\hat{\psi})$ is not a diagonal matrix , we are not able to give an expression for the adjusted profile likelihood for ρ and σ^2. This is not a problem in terms of inference, since we have the standard errors for hypothesis testing or we can use the score test which is based on the likelihood score and its derivatives. Another approach we

may take is to extend the method proposed by Stafford (1996) and used to obtain (3.60). We assume that σ^2 is equal to its adjusted maximum likelihood estimate and approximate W^* by a diagonal matrix so that the adjusted score for ρ is

$$\tilde{U}_\rho = 1.0349\left(\frac{\partial \ell_{\mathrm{p}}}{\partial \rho} - m_1\right) + 0.00026\left(\frac{\partial \ell_{\mathrm{p}}}{\partial \sigma^2} - m_2\right) \tag{3.67}$$

$$\approx k\left(\frac{\partial \ell_{\mathrm{p}}}{\partial \rho} + \mathrm{tr}(PWD^{-1})\right) \tag{3.68}$$

where k is a scaling constant to be determined so that (3.68) gives the correct adjusted maximum likelihood estimate ($\hat{\rho} = 0.336$) and the correct standard error (0.114); (3.68) satisfies the first property for all k, and the second is satisfied by taking $k = 1.0345$ (the change in the value of k from (3.67) to (3.68) occurs because we are assuming that the adjustment for ρ is independent of the score function for σ^2).

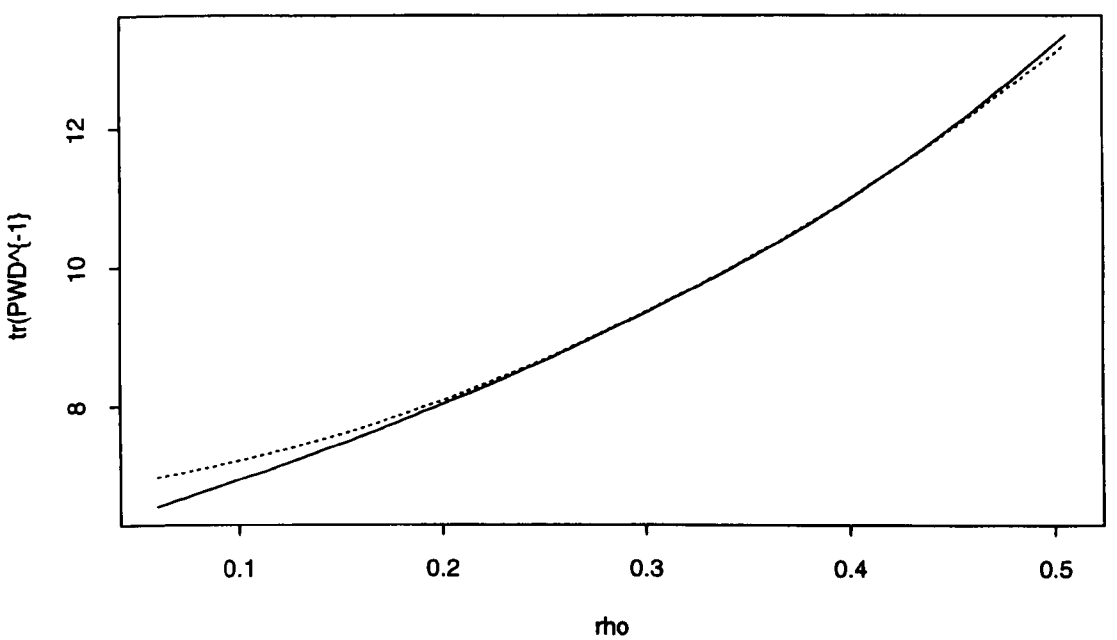

Figure 3.2: *Plot of* $tr(PWD^{-1})$ —– *and the quadratic function* $q(\rho)$ - - - *with the same first two derivatives at* $\rho = \hat{\rho}$.

We need to integrate (3.68) to obtain an adjusted profile log-likelihood for ρ. The integral of $\mathrm{tr}(PWD^{-1})$ does not exist as a standard function. However, a plot of $\mathrm{tr}(PWD^{-1})$ against ρ (see Figure 3.2) suggests that a quadratic approximation would be appropriate. The quadratic function $q(\rho)$ is chosen so that $\mathrm{tr}(PWD^{-1})$ and $q(\rho)$ have the same two first derivatives at the adjusted estimate of ρ. The adjusted profile log-likelihood is now easily found as

$$\ell_{\mathrm{ap}}(\rho) = 1.0345\ell_{\mathrm{p}}(\rho, \hat{\sigma}^2) - 54.151 + 6.754\rho + 1.3705\rho^2 + 6.507\rho^3 \qquad (3.69)$$

where we have adjusted the function so that the maximum value of $\ell_{\mathrm{ap}}(\rho)$ is zero. Figure 3.3 shows the profile and adjusted profile log-likelihood for ρ as given in (3.63) and (3.69) where both likelihoods have been adjusted so that their maximum is zero.

Remark 1 We obtain again the usual large adjustment for the estimate of σ^2, since the degrees of freedom of the residual sum of squares are reduced from 36 to 24. There is also a large adjustment for ρ. We performed a Monte Carlo simulation study in order to verify whether this adjustment was desirable. We generated 100 sets of data from (3.61) for the Kempton model. The X matrix and parameter values were those found for the Jenkyn et al. (1979) data. Several values of ρ were taken (between -0.3 and 0.3) and the value of σ^2 was taken as 0.03 (the adjusted estimate of σ^2 was 0.035). When ρ was taken as 0.3, the average of the adjusted estimates of ρ over the 100 simulations was 0.299 with standard error 0.110, while the average value of the unadjusted estimates was 0.206 with standard error 0.100. In all other cases the average of the adjusted estimates of ρ was very close to the true value.

Remark 2 The increase in the estimate of σ^2 increases the standard errors of the estimated effects, but the increase in $\hat{\rho}$ changes the estimated effects themselves. Table 3.1 gives the estimated monoculture means (estimated treatment yields under the assumption that the treatment is applied to the entire field (Kempton, 1982)).

Remark 3 The adjusted score in (3.68) and the adjusted profile log-likelihood in (3.69) have been derived for the Kempton model as applied to Jenkyn et al. (1979)

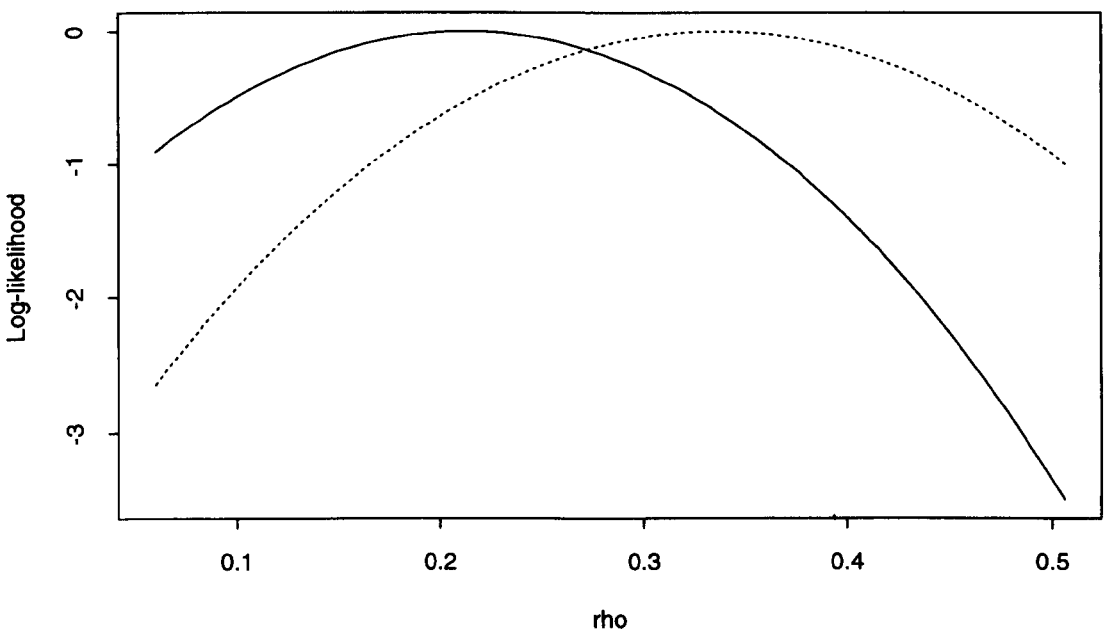

Figure 3.3: *Profile likelihood —– and adjusted profile likelihood - - - for Kempton example.*

	Treatments			
ρ	1	2	3	4
0	5.31	5.87	6.09	5.94
0.231	5.11	5.89	6.17	6.03
0.336	4.97	5.91	6.23	6.08

Table 3.1: *Estimated monoculture means for various values of ρ.*

data. However, the method is more general. The derivation of (3.68) depends first on been able to replace any "other" parameters of interest by their adjusted estimates and second on being able to approximate W^* by a diagonal matrix. The resulting score will give the correct adjusted estimate but may need rescaling to give the same variance.

Two other data sets in which the Kempton model has been used have been analysed, though the change in the competition parameter ρ was not so large. In Kempton (1982) the competition parameter was estimated as $\hat{\rho} = -0.294$, while with (3.36) we obtain $\hat{\rho} = -0.281$. In Besag and Kempton (1986) $\hat{\rho}$ changed from -0.372 to -0.351. A possible reason why the change in the competition parameter is so large in our first data set may be the fact it is trend and not competition that is present in the data, and both effects are being confounded.

Chapter 4

Smoothing in the presence of a competition effect

4.1 Introduction

Inter-plot competition models, such as models (3.59) and (3.66), aim to account for competition effects which may bias the estimates of the variety means. These models were first introduced by Draper and Guttman (1980), Kempton (1982) and Besag and Kempton (1986). These models assume *"the experiment is one in which competition is locally much more important than differences in fertility"* (Besag and Kempton, 1986), and therefore, blocks need to be small. Kempton (1984) pointed out the difficulty in seeing *"how a nearest neighbour analysis could be included in (3.66) to remove more local trends in fertility"*. When these trends in fertility are present, they may affect the magnitude of the competition parameter and it may be difficult to separate the positive association due to fertility trends and competition. The aim of this chapter is to give a possible solution to these problems by accounting for the fertility trends by means of linear smoothers.

The combination of a smooth trend together with a competition effect may be seen as the extension of two classical models in the analysis of agricultural trials: the competition models given by Kempton (1982) and by Draper and Guttman (1980) accounted

for spatial variation using blocks. We extend these models by accounting for the global trend in a more continuous fashion. In section 4.2.1, the semi-parametric model introduced by Green et al. (1985) is extended to account for a possible local competition effect within a small neighbourhood. In section 4.2.2 we expand these models to any number of smooth terms using the model of Besag and Kempton (1986). The model proposed by Draper and Guttman (1980) is combined with an additive model in section 4.3. We also extend Kempton's model using the mixed model decomposition of a cubic smoothing spline (section 4.4); this approach allows us to account for fertility trends and competition effects in a fully parametric model, and therefore, smoothing and competition parameters may be estimated simultaneously (section 4.5).

4.2 Competition between plots

4.2.1 Models with one smooth term

Besag and Kempton (1986) proposed a model which incorporated inter-plot competition together with treatment and block effects. Our model is inspired by this model, but blocks are replaced with a smooth function of plot position in the field. This model is written as:

$$y = X\beta + f(t) + \rho W y + \epsilon \tag{4.1}$$

where X is the $n \times p$ treatment design matrix, $f(t)$ is an unknown smooth function of plot position in the field and W is a $n \times n$ weight matrix with off-diagonal elements $w_{i,i-1} = w_{i,i+1} = 1/2$ and zero otherwise. We make the standard assumption about the distribution of the errors,

$$\epsilon \sim MVN\left(0, \sigma^2 I\right).$$

We will use the adjusted profile likelihood of Chapter 3 to estimate the parameters in this non-linear model. However, $f(t)$ is a non-parametric term, and so it will be necessary to rewrite $f(t)$ in parametric form to allow the maximum likelihood estimates to be found. If the smoother used is a cubic smoothing spline, it is possible to find a mixed model formulation which fits a non-parametric smoother in a parametric

setting (see section 4.4). However, this is not the case for other smoothers, e.g., loess. Following the idea used in Speckman (1988), we suppose that $f(t)$ can be parameterized as: $f(t) = T\gamma$ where γ is an additional vector of parameters. The idea is to derive estimates of β and γ in terms of X and T and then replace any occurrence of $T(T'T)^{-1}T'$ by the smoother S, as suggested by Green et al. (1985). Model (4.1) can now be written as:

$$y = L\delta + \rho Wy + \epsilon \tag{4.2}$$

where $L = [X : T]$ and $\delta' = (\beta', \gamma')$. This converts (4.1) to the Besag and Kempton model. We write (4.2) as

$$\begin{aligned} (I - \rho W)y &= L\delta + \epsilon \\ y &= EL\delta + E\epsilon \end{aligned} \tag{4.3}$$

where

$$D = I - \rho W \quad \text{and} \quad E = D^{-1}. \tag{4.4}$$

Thus,

$$y \sim MVN(EL\delta, EE'\sigma^2).$$

We use the profile likelihood to estimate $\theta = (\delta, \sigma^2, \rho)$. We take $\psi = (\sigma^2, \rho)$ as parameters of interest and δ as nuisance parameters. Then, the maximum likelihood estimates of β and γ are obtained as functions of ψ (we will denote these estimates by $\hat{\delta}_\psi = (\hat{\beta}_\psi, \hat{\gamma}_\psi)$) and substituted in the log-likelihood. The likelihood for θ is given by:

$$\mathrm{L}(\delta, \sigma^2, \rho \mid y) \propto (\sigma^2)^{-\frac{n}{2}} |D| \exp\left(-\frac{1}{2\sigma^2} R_\rho\right) \tag{4.5}$$

where

$$R_\rho = (Dy - L\delta)'(Dy - L\delta). \tag{4.6}$$

The log-likelihood is (omitting additive constants):

$$\ell(\beta, \gamma, \rho, \sigma^2 \mid y) = -\frac{n}{2} \log \sigma^2 + \log |D| - \frac{1}{2\sigma^2} R_\rho. \tag{4.7}$$

In the following lemma, we give the expressions for G, H, and Q (defined in section 3.2) for model (4.1).

Lemma 4.1 *The matrices G, H and Q given in (3.8), (3.9) and (3.10) are*

$$G = \sigma^2 \begin{pmatrix} U & -UX'T(T'T)^{-1} \\ -(T'T)^{-1}T'XU & (T'T)^{-1} + (T'T)^{-1}T'XUX'T(T'T)^{-1} \end{pmatrix} \quad (4.8)$$

$$H = \frac{1}{\sigma^2} D'MD \quad (4.9)$$

$$Q = \frac{1}{\sigma^2} D'(I - M)D \quad (4.10)$$

where

$$U = (X'(I - S)X)^{-1} \quad (4.11)$$

$$M = S + (I - S)XUX'(I - S) \quad (4.12)$$

and S is the smoother matrix which replaces $T(T'T)^{-1}T'$.

Proof From the definition of G given in (3.8),

$$G = \sigma^2 (L'L)^{-1}$$

where

$$L'L = \begin{pmatrix} X'X & X'T \\ T'X & T'T \end{pmatrix}.$$

Using now the formula for the inverse of a partitioned matrix (see Schott, 1997):

$$\begin{pmatrix} A & B \\ B' & C \end{pmatrix}^{-1} = \begin{pmatrix} I & 0 \\ -C^{-1}B' & C^{-1} \end{pmatrix} \begin{pmatrix} (A - BC^{-1}B')^{-1} & -(A - BC^{-1}B')^{-1}BC^{-1} \\ 0 & I \end{pmatrix}, \quad (4.13)$$

we find that for $A = X'X$, $B = X'T$ and $C = T'T$

$$(A - BC^{-1}B')^{-1} = (X'X - X'T(T'T)^{-1}T'X)^{-1} = (X'(I - S)X)^{-1} = U.$$

The expression for G given in (4.8) follows from application of (4.13).

The definition of H given in (3.9) yields

$$\begin{aligned} H &= \frac{1}{\sigma^4} D'LGL'D \\ &= \frac{1}{\sigma^4} D' \begin{bmatrix} X & : & T \end{bmatrix} G \begin{bmatrix} X' \\ T' \end{bmatrix} D \end{aligned}$$

$$= \frac{1}{\sigma^2} D' \left[(I-S)XU \;:\; T(T'T)^{-1} - (I-S)XUX'T(T'T)^{-1} \right] \begin{bmatrix} X' \\ T' \end{bmatrix} D$$

$$= \frac{1}{\sigma^2} D'[S + (I-S)X'UX(I-S)]D$$

$$= \frac{1}{\sigma^2} D'MD \quad \text{from (4.12).}$$

The proof of (4.10) is immediate from the definition of Q given in (3.10) and (4.9).■

Remark The expression for Q given in (3.10) is symmetric and satisfies $Q\Sigma Q = Q$. In the case of model (4.1), Q is symmetric if the smoother is symmetric. Further,

$$Q'\Sigma Q = \frac{1}{\sigma^2} D'(I-P)D \tag{4.14}$$

where

$$P = M + M' - M'M. \tag{4.15}$$

Thus, $Q'\Sigma Q \neq Q$ unless the smoother matrix S is idempotent (which is the case for projection smoothers).

In the following lemma we give expressions for the estimates of the fixed and smooth terms in model (4.1).

Lemma 4.2 *The estimated treatment and smooth effects in equation (4.1) are*

$$\hat{\beta}_\rho = FDy \tag{4.16}$$

$$\hat{f}_\rho = S(I-XF)Dy \tag{4.17}$$

where

$$F = UX'(I-S) \tag{4.18}$$

Proof Equation (3.12) gives the expression for the estimates of the fixed terms in the model. For model (4.3) we have

$$\hat{\delta} = (L'L)^{-1}L'Dy \tag{4.19}$$

which yields

$$\hat{\beta}_\rho = UX'(I-S)Dy = FDy$$

$$\hat{f}_\rho = T\hat{\gamma}_\rho = S(I-XF)Dy$$

by means of (4.8) and (4.18). ■

Remark The estimates of β and f given in (4.16) and (4.17) are the standard estimates of a semiparametric model but for the adjusted response Dy, and the partial derivatives of the log-likelihood (4.5) give rise to the normal equations for a semiparametric model with one smooth term and response Dy.

Now, we can give an expression for the profile log-likelihood for model (4.3)

$$\ell_p(\psi) = -\frac{n}{2}\log\sigma^2 + \log|D| - \frac{1}{2\sigma^2}y'D'(I-P)Dy, \tag{4.20}$$

where P is given by (4.15) and (4.12). We note that we cannot apply the results of chapter 3 to obtain the adjustments of the score function since $Q\Sigma Q \neq Q$. However, a routine application of the definitions in (3.1) to (3.4) to the profile likelihood (4.20) yields the following theorem which gives the bias and variance adjustments to the score function.

Theorem 4.1 *The bias and variance adjustments of the profile log-likelihood for model (4.1) are given by*

$$m'(\sigma^2,\rho) = \left(-\frac{1}{2\sigma^2}tr(P), tr(A)\right) \tag{4.21}$$

$$W_1(\sigma^2,\sigma^2) = W_2(\sigma^2,\sigma^2) = \frac{1}{2\sigma^4}tr(I-P) \approx \frac{n-p-q}{2\sigma^4} \tag{4.22}$$

$$W_1(\sigma^2,\rho) = W_1(\rho,\sigma^2) = W_2(\sigma^2,\rho) = W_2(\rho,\sigma^2) = -\frac{1}{\sigma^2}tr(B) \tag{4.23}$$

$$W_1(\rho,\rho) = tr(B^2) + tr(B'B) + \frac{1}{\sigma^2}(X\beta+f)'B'B(X\beta+f) \tag{4.24}$$

$$W_2(\rho,\rho) = W_1(\rho,\rho) + tr(AB). \tag{4.25}$$

where

$$A = -PWE \tag{4.26}$$

$$B = -(I-P)WE \tag{4.27}$$

and P is given in (4.15) and (4.12).

Proof The expressions for the bias and variance adjustments follow from the definition of m, W_1 and W_2 given in (3.1) to (3.3) and assuming that the bias in the nonparametric part of the model can be ignored. ■

Remark 1 It follows from (4.20) and (4.21) that the estimate of σ^2 from the adjusted profile log-likelihood score is

$$\hat{\sigma}^2 = \frac{y'D'(I-P)Dy}{\text{tr}(I-P)} \tag{4.28}$$

which is the estimate of σ^2 given on (2.52) on the adjusted response Dy.

Remark 2 Expressions (4.21) to (4.25) are similar to the bias and variance adjustments for model (3.61) given in page 64, although in Theorem 4.1, P is not idempotent. In the presence of competition, additive models are again the natural extension of linear models.

Remark 3 The bias and variance adjustment depend on the smoother matrix S and therefore on the smoothing parameter λ. In section 4.5 we will present different approaches to the estimation of the smoothing parameter in the presence of competition.

4.2.2 Extension to two or more smooth terms

The number of smooth terms in the analysis of agricultural data is usually one (if we use a two-dimensional smooth surface to explain the spatial variation) or two (when an additive model is preferred). Thus, it would be of interest to extend the previous model with a single smooth term to a more general situation.

We extend model (4.1) to:

$$y = X\beta + f_1(z_1) + ... + f_q(z_q) + \rho Wy + \epsilon \tag{4.29}$$

where, $f_1, \ldots, f_q$ are q unknown smooth functions and $z_1, \ldots, z_q$ are the q covariates against which we want to smooth.

Then, (4.3) becomes:

$$y = EX\beta + Ef_1 + ... + Ef_q + E\epsilon$$

Again, assuming that each smooth term can be parameterised as $T_i\gamma_i$, $i = 1 \ldots q$:

$$y \sim MVN(EX\beta + E(T_1\gamma_1 + \ldots + T_q\gamma_q), EE'\sigma^2).$$

The likelihood for β, $\gamma_1, \ldots, \gamma_q$, ρ and σ^2 has the same form as (4.5), where

$$R_\rho = (Dy - X\beta - T_1\gamma_1 - \ldots - T_q\gamma_q)'(Dy - X\beta - T_1\gamma_1 - \ldots - T_q\gamma_q). \qquad (4.30)$$

Taking derivatives with respect to β, $\gamma_1, \ldots, \gamma_q$, we find :

$$\begin{aligned}
\frac{\partial l}{\partial \beta} &= -\frac{2}{\sigma^2}X'(Dy - X\beta - T_1\gamma_1 - \ldots - T_q\gamma_q) \\
\frac{\partial l}{\partial \gamma_1} &= -\frac{2}{\sigma^2}T_1'(Dy - X\beta - T_1\gamma_1 - \ldots - T_q\gamma_q) \\
&\vdots \\
\frac{\partial l}{\partial \gamma_q} &= -\frac{2}{\sigma^2}T_q'(Dy - X\beta - T_1\gamma_1 - \ldots - T_q\gamma_q).
\end{aligned}$$

These expressions give rise to the normal equations of a semi-parametric model with q smooth terms and response Dy. In (2.41) and (2.43) we showed that:

$$\hat{\beta}_\rho = U_qX'(I - M_q)Dy \qquad (4.31)$$

$$\hat{f}_1 + ... + \hat{f}_q = M_q(I - XF_q)Dy \qquad (4.32)$$

where

$$F_q = U_qX'(I - M_q) \quad \text{and} \quad U_q = (X'(I - M_q)X)^{-1}$$

and M_q is the centred hat matrix of an additive model with q smooth terms. Thus, the profile-likelihood is

$$\ell_p(\psi) = -\frac{n}{2}\log\sigma^2 + \log|D| - \frac{1}{2\sigma^2}y'D'(I - P_q)Dy. \qquad (4.33)$$

where

$$P_q = M_q + M_q' - M_q'M_q,$$

and the bias and variance adjustments are similar to those given in Theorem 4.1, substituting P by P_q.

4.3 Competition between neighbouring treatments

Besag and Kempton (1986) suggested that for some crops (such as cereals), *"a model which allows varietal competition effects to be individually specified, is useful not only for yield adjustment but also to aid biological interpretation"*. Pearce (1957) introduced this model which assumes that the treatment applied to a particular plot has an effect on the treatments applied to the neighbouring plots:

$$y = T\gamma + X\beta + WX\Phi + \epsilon \tag{4.34}$$

where T, X and W are defined as in the previous section and Φ is a vector of centred treatment neighbour effects. Of course, if the number of treatments is large (as in the case of a variety trial) the number of parameters to be estimated becomes extreme. Draper and Guttman (1980) use $\Phi = \rho\beta$ and so avoid the problem of a large number of parameters. In this model, there is a common competition coefficient for all treatments; hence,

$$Y = JX\beta + T\gamma + \epsilon \tag{4.35}$$

or

$$Y = JX\beta + f(z) + \epsilon \tag{4.36}$$

where $J = I + \rho W$ and $f(z)$ is a smooth function of plot position in the field. Besag and Kempton (1986) proposed the use of least squares estimation, iterating between the estimation of β and ρ after adjusting for the block effect. Here, we adopt the same approach as in the case of competition between neighbouring plots: for convenience, we suppose that $f(z)$ may be parameterised as $T\gamma$ and use the profile likelihood to estimate ρ and σ^2. Hence, model (4.36) becomes

$$y = L\delta + \epsilon \tag{4.37}$$

where $L = [JX : T]$, $\delta' = (\beta', \gamma')$. Hence, the log-likelihood for $\theta = (\delta, \sigma^2, \rho)$ is given by:

$$\mathrm{L}(\delta, \sigma^2, \rho \mid y) = -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}R_\rho \tag{4.38}$$

where

$$R_\rho = (y - L\delta)'(y - L\delta). \tag{4.39}$$

Again, we consider δ as a vector of nuisance parameters and $\psi = (\sigma^2, \rho)$ as parameters of interest and use profile likelihood. The estimates of θ and ψ depend on the matrices G, H and Q that we give in the following lemma.

Lemma 4.3 *For model (4.37), the matrices G, H and Q defined in (3.8), (3.9) and (3.10) are*

$$G = \sigma^2 \begin{pmatrix} U_J & -U_J X'J'T(T'T)^{-1} \\ -(T'T)^{-1}T'JXU_J & (T'T)^{-1} + (T'T)^{-1}T'JXU_JX'J'T(T'T)^{-1} \end{pmatrix} \quad (4.40)$$

$$H = \frac{1}{\sigma^2}M \quad (4.41)$$

$$Q = \frac{1}{\sigma^2}(I - M) \quad (4.42)$$

where

$$U_J = (X'J'(I-S)JX)^{-1} \quad (4.43)$$

$$M_J = S + (I-S)JXU_JX'J'(I-S), \quad (4.44)$$

S being the smoother matrix which replaces $T'(T'T)^{-1}T$.

Proof From the definition of G given in (3.8),

$$G = \sigma^2(L'L)^{-1}$$

and

$$L'L = \begin{pmatrix} X'J'JX & X'J'T \\ T'JX & T'T \end{pmatrix} \quad J = I + \rho W.$$

G follows from the expression for the inverse of a partitioned matrix given in (4.13). (4.41) and (4.42) follow from the definition of H and Q and (4.40). ■

Lemma 4.4 *The estimates of* β *and* f *in model (4.36) are*

$$\hat{\beta}_\rho = F_J y \quad (4.45)$$

$$\hat{f}_\rho = S(I - JXF_J)y \quad (4.46)$$

where

$$F_J = U_J X'J(I-S). \quad (4.47)$$

Proof From (3.12) we have

$$\hat{\delta} = (L'L)^{-1}L'y$$

which yields

$$\begin{aligned} \hat{\beta}_\rho &= U_J X' J'(I - S)y = F_J y \\ \hat{f}_\rho = T\hat{\gamma}_\rho &= S(I - JXF_J)y \end{aligned}$$

from (4.40) and (4.47). ∎

Remark (4.45) and (4.46) are the estimates of the linear and smooth part of a semi-parametric model (2.40) where the design matrix is JX.

Model (4.37) is a special case of the model presented in section 3.5.2 (Models with independent and identically distributed errors), but in this case Q is not idempotent and hence, the profile log-likelihood for model (4.37) is

$$\ell_p(\psi) = -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}y'(I - P_J)y \tag{4.48}$$

where $P_J = M_J + M_J' - M_J'M_J$ and M_J is given in (4.44).

Theorem 4.2 *The bias-adjusted profile log-likelihood for model (4.36) is*

$$\ell_{ap} = -\frac{n-p-q}{2}\log\sigma^2 - \frac{1}{2\sigma^2}y'(I - P_J)y \tag{4.49}$$

and the variance adjustment for the profile log-likelihood score is given by

$$W_1(\sigma^2,\sigma^2) = W_2(\sigma^2,\sigma^2) = \frac{1}{2\sigma^4}tr(I - P_J) \approx \frac{n-p-q}{2\sigma^4} \tag{4.50}$$

$$W_1(\sigma^2,\rho) = W_1(\rho,\sigma^2) = W_2(\sigma^2,\rho) = W_2(\rho,\sigma^2) = 0 \tag{4.51}$$

$$W_1(\rho,\rho) \approx tr(U_J X'W'(I - M_J)WX)) + \beta X'W'(I - M_J)WX\beta \tag{4.52}$$

$$W_2(\rho,\rho) \approx \beta X'W'(I - M_J)WX\beta. \tag{4.53}$$

where p is the rank of X and q is the trace of S.

Proof The bias adjustment of the profile log-likelihood score for model (4.36) is

$$m'(\rho,\sigma^2) = \left(0, -\frac{1}{2\sigma^2}\mathrm{tr}(P_J)\right) \approx \left(0, -\frac{p+q}{2\sigma^2}\right)$$

by (3.1), the fact that the covariance matrix does not depend on ρ and assuming again $Sf \approx f$. For the calculation of the variance adjustment, we assume, for simplicity that

$$y'(I - P_J)y \approx y(I - M_J)y.$$

The result follows from application of equations (3.53) to (3.58). ■

Remark The variance adjustment $W = W_2' W_1^{-1}$ has the form

$$W(\rho, \sigma^2) = \begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix}$$

and $a \approx 1$.

In the next section we take a different approach to smoothing (when the smoother used is a cubic smoothing spline) in the presence of competition, based on the relationship between smoothing spline models and mixed-effects models.

4.4 Smoothing splines as mixed models and competition

Wang (1998b) and Wang (1998a) have shown that cubic smoothing splines may be formulated as mixed models and that inference can be carried out by generalised maximum likelihood. Verbyla et al. (1998) have used this approach to analyse designed experiments. We use this approach to estimate the competition and smoothing parameter simultaneously and we show that the estimates of the fixed effects and smooth term are equivalent to those given in (4.16) and (4.17).

First, we will introduce some notation and give some identities. The next section is based on Green and Silverman (1994) and Verbyla et al. (1998).

4.4.1 Notation

Given the non-parametric regression model

$$y = f(z) + \epsilon, \quad \text{and} \quad \epsilon \sim \mathcal{N}(0, \sigma^2\Sigma), \tag{4.54}$$

let $h_i = z_{i+1} - z_i$ for $i = 1, \ldots, n-1$: without loss of generality we will assume that z has been centred, i.e. we take z as $z - \frac{1}{n}\sum_{i=1}^{n} z_i$. We define two matrices Δ and G; these are the matrices Q and R in Green and Silverman (1994), p 12. Let Δ be a $n \times (n-2)$ matrix with entries

$$\left.\begin{array}{ll} \Delta_{ij} = 0 & \text{for } \mid i-j \mid \geq 2 \\ \Delta_{jj} = h_j^{-1} & \\ \Delta_{j+1,j} = -h_j^{-1} - h_{j+1}^{-1} & \text{for } j = 1, \ldots, n-2 \\ \Delta_{j+2,j} = h_{j+1}^{-1} & \end{array}\right\} \tag{4.55}$$

and G be a symmetric $(n-2) \times (n-2)$ matrix with entries

$$\begin{array}{ll} G_{ii} = \frac{1}{3}(h_i + h_{i+1}) & \text{for } i = 1, \ldots, n-2 \\ G_{i,i+1} = G_{i+1,i} = \frac{1}{6}h_{i+1} & \text{for } i = 2, \ldots, n-3 \\ G_{ij} = 0 & \text{for } \mid i-j \mid \geq 2 \end{array}.$$

Green and Silverman (1994) p 41, showed (using penalised maximum likelihood) that the estimate of f in (4.54) is given by

$$\hat{f} = \underbrace{(\Sigma^{-1} + \lambda\Delta G^{-1}\Delta')^{-1}\Sigma^{-1}}_{S^*} y \tag{4.56}$$

where λ is the smoothing parameter and Δ and G defined as above. For simplicity and to follow the assumptions in the previous chapters, we will assume that $\Sigma = I$ in the rest of the section. However, the results will apply in the general case. Thus, we have

$$\hat{f} = \underbrace{(I + \lambda\Delta G^{-1}\Delta')^{-1}}_{S^*} y \quad \text{Green and Silverman (1994) p19.} \tag{4.57}$$

The equivalence between the mixed model and the additive approach depends on some identities satisfied by Δ and S^*. We give them in the following lemma.

Lemma 4.5 *With Δ and S^* defined in (4.55) and (4.57) respectively, we have for the centred predictor z*

$$1'\Delta = \underline{0}' \tag{4.58}$$

$$z'\Delta = \underline{0}' \tag{4.59}$$

$$S^* = I - \Delta(\Delta'\Delta)^{-1}\Delta' + N_s\left(N_s'N_s + \lambda G^{-1}\right)^{-1}N_s' \tag{4.60}$$

where

$$N_s = \Delta(\Delta'\Delta)^{-1}. \tag{4.61}$$

Proof $1'\Delta$ is a vector whose elements are the sum of each of the columns of Δ, thus, the jth element of $1'\Delta$ is

$$(1'\Delta)_j = h_j^{-1} - h_j^{-1} - h_{j+1}^{-1} + h_{j+1}^{-1} = 0 \text{ from (4.55).}$$

The elements of $z'\Delta$ are

$$\begin{aligned}(z'\Delta)_j &= z_j\Delta_{jj} + z_{j+1}\Delta_{j+1,j} + z_{j+2}\Delta_{j+2,j}\\ &= \frac{z_j}{h_j} - \frac{z_{j+1}}{h_j} - \frac{z_{j+1}}{h_{j+1}} + \frac{z_{j+2}}{h_{j+1}} \quad \text{from (4.55)}\\ &= -\frac{h_j}{h_j} + \frac{h_{j+1}}{h_{j+1}} \quad \text{from the definition of } h_i\\ &= 1 - 1 = 0.\end{aligned}$$

To prove (4.60) we use the formula for the inverse of the sum of matrices given in (2.19)

$$\begin{aligned}S^* &= I - \Delta(\Delta'\Delta + \lambda^{-1}G)^{-1}\Delta'\\ &\quad \text{we apply (2.19) again to obtain}\\ &= I - \Delta\left[(\Delta'\Delta)^{-1} - (\Delta'\Delta)^{-1}\left((\Delta'\Delta)^{-1} + \lambda G^{-1}\right)^{-1}(\Delta'\Delta)^{-1}\right]\Delta'\\ &= I - \Delta(\Delta'\Delta)^{-1}\Delta' + \Delta(\Delta'\Delta)^{-1}\left((\Delta'\Delta)^{-1} + \lambda G^{-1}\right)^{-1}(\Delta'\Delta)^{-1}\Delta'\end{aligned}$$

which yields (4.60) from (4.61). ■

Remark From (4.58) and (4.59) we can immediately show the standard results that

the smoother matrix of a cubic smoothing spline satisfies (Hastie and Tibshirani, 1990)

$$z'S^* = z' \quad S^*z = z \tag{4.62}$$

$$1'S^* = 1' \quad S^*1 = 1$$

and as a consequence the centred smoother matrix for a smoothing spline is

$$S = (I - 11'/n)S^* = S^* - 11'/n. \tag{4.63}$$

Verbyla et al. (1998) have shown that the non-parametric part ($f(z)$) of a model of the form

$$y = f(z) + \epsilon$$

may be written as

$$f(z) = X_s\beta_s + N_s u_s \tag{4.64}$$

where u_s is a random vector, $u_s \sim \mathcal{N}(0, \sigma_s^2 G)$, X_s is a $n \times 2$ matrix whose columns are a vector of ones and the vector z. Hence, model (4.1) for competition between neighbouring plots may be written as

$$y = X\beta + X_s\beta_s + N_s u_s + \rho W y + \epsilon \tag{4.65}$$

and so

$$y = EX\beta + EX_s\beta_s + EN_s u_s + E\epsilon \tag{4.66}$$

where E is given by equation (4.4). To avoid identifiability problems, we will assume that 1 is in the column span of X and $X_s = z$ the column vector of the centred predictor, and therefore, $f(z)$ as given in (4.64) is a centred smoother. In a model like (4.65) with $\rho = 0$, Wang (1998b) and Verbyla et al. (1998) use the mixed model equations (Henderson et al., 1959) to give estimates of the fixed and random effects and REML to give estimates of the variance components. When ρ is present in the model, it is not possible to use REML to estimate the variance components and ρ (since ρ is present in the mean and the variance simultaneously). Our approach will be to estimate the random and fixed effects in the same way as Wang (1998b) and Verbyla et al. (1998), then substitute the estimate of the fixed effect into the density of y and use the profile likelihood for ρ and the other variance parameters.

4.4.2 Estimation of the random and fixed effects

Given $\epsilon \sim \mathcal{N}(0, \sigma^2 I)$ and $u_s \sim \mathcal{N}(0, \sigma_s^2 G)$, the mixed model equations (Henderson et al., 1959) for model (4.66) are

$$\underbrace{\begin{bmatrix} X'X & X'N_s & X'X_s \\ N_s'X & N_s'N_s + \lambda G^{-1} & 0 \\ X_s'X & 0 & X_s'X_s \end{bmatrix}}_{L} \begin{bmatrix} \hat{\beta} \\ \tilde{u}_s \\ \hat{\beta}_s \end{bmatrix} = \begin{bmatrix} X'Dy \\ N_s'Dy \\ X_s'Dy \end{bmatrix} \tag{4.67}$$

where D is defined in (4.4) and

$$\lambda = \sigma^2/\sigma_s^2. \tag{4.68}$$

Note that $L_{2,3} = L_{3,2} = 0$ since $X_s = z$ and by (4.61) and (4.59). This set of equations is the same as the ordinary mixed model equations for the adjusted response Dy. In the following lemma we give an expression for the centred smoother matrix S in terms of X_s, N_s and λ.

Lemma 4.6 *The centred smoother matrix as defined in (4.63) can be written as*

$$S = [N_s, X_s] \begin{bmatrix} N_s'N_s + \lambda G^{-1} & 0 \\ 0 & X_s'X_s \end{bmatrix}^{-1} \begin{bmatrix} N_s' \\ X_s' \end{bmatrix}. \tag{4.69}$$

Proof From (4.60) and (4.63) we have

$$S = I - \Delta(\Delta'\Delta)^{-1}\Delta' + N_s \left(N_s'N_s + \lambda G^{-1}\right)^{-1} N_s' - 11'/n.$$

Using the identity $\Delta(\Delta'\Delta)^{-1}\Delta' = I - X_s(X_s'X_s)^{-1}X_s' - 11'/n$ (Khatri, 1966), for Δ defined in (4.55) and X_s the vector of the centred covariate, we find

$$S = X_s(X_s'X_s)^{-1}X_s' + N_s \left(N_s'N_s + \lambda G^{-1}\right)^{-1} N_s'$$

which is (4.69). ■

We give the expression for the estimates of the fixed and random effects in the following theorem.

Theorem 4.3 *For given λ and ρ the estimates of the fixed and random effects in model (4.66) are*

$$\hat{\beta} = (X'(I-S)X)^{-1}X'(I-S)Dy \tag{4.70}$$

$$\hat{\beta}_s = (X_s'X_s)^{-1}X_s'(Dy - X\hat{\beta}) \tag{4.71}$$

$$\tilde{u}_s = (N_s'N_s + \lambda G^{-1})^{-1}N_s'(Dy - X\hat{\beta}) \tag{4.72}$$

and

$$\hat{f} = S(Dy - \hat{\beta}). \tag{4.73}$$

Proof From (4.67) we have

$$\begin{bmatrix} \hat{\beta} \\ \tilde{u}_s \\ \hat{\beta}_s \end{bmatrix} = L^{-1} \begin{bmatrix} X' \\ N_s' \\ X_s' \end{bmatrix} Dy.$$

In order to calculate L^{-1} we use (4.13) where

$$\begin{aligned} A &= X'X \\ B &= X'[N_s, X_s] \\ C &= \begin{bmatrix} N_s'N_s + \lambda G^{-1} & 0 \\ 0 & X_s'X_s \end{bmatrix} \end{aligned}$$

From (4.13)

$$\hat{\beta} = (A - BC^{-1}B')^{-1}X' \left(I - [N_s, X_s]C^{-1} \begin{bmatrix} N_s' \\ X_s' \end{bmatrix} \right) Dy \tag{4.74}$$

$$\begin{bmatrix} \tilde{u}_s \\ \hat{\beta}_s \end{bmatrix} = C^{-1} \begin{bmatrix} N_s' \\ X_s' \end{bmatrix} \left(Dy - X\hat{\beta} \right), \tag{4.75}$$

but

$$\begin{aligned} A - B'C^{-1}B &= X' \left(I - [N_s, X_s]\, C^{-1} \begin{bmatrix} N_s' \\ X_s' \end{bmatrix} \right) X \\ &= X'(I-S)X \quad \text{by (4.69).} \end{aligned} \tag{4.76}$$

Substituting (4.76) in (4.74) and (4.75) we obtain (4.70), (4.71) and (4.72). (4.73) follows from the definition of f given in (4.64) and (4.74) and (4.75). ■

Remark For fixed values of λ and ρ, taking f as a fixed effect (section 4.2.1) or random (4.64) yields the same estimates of β in (4.16) and f in (4.17).

In the next section we use again the method of the profile likelihood to estimate σ^2, σ_s^2 (or equivalently λ) and ρ.

4.4.3 Estimation of the variance components and competition parameter

From (4.66) the distribution of y is

$$y \sim \mathcal{N}(EX\beta + EX_s\beta_s, \underbrace{\sigma^2 E(I + \lambda^{-1} N_s G N_s')E'}_{\Sigma}) \tag{4.77}$$

where λ is given by (4.68). We are interested in using the profile likelihood for the model

$$y = EX\beta + EX_s\beta_s + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \Sigma) \tag{4.78}$$

where Σ is given in (4.77). We will take β and β_s as nuisance parameters and $\psi = (\sigma^2, \sigma_s^2, \rho)$ or equivalently $\psi = (\sigma^2, \lambda, \rho)$ as parameters of interest. The results that follow in the next sections will be a generalisation of those given in section 3.5.3 (in that section we assumed $\Sigma = \sigma^2 I$).

Lemma 4.7 *The covariance matrix Σ given in (4.77) satisfies the following identities*

$$\Sigma^{-1} = \frac{1}{\sigma^2} D'(I - S + X_s(X_s'X_s)^{-1}X_s')D \tag{4.79}$$

$$\Sigma^{-1} EX_s = \frac{1}{\sigma^2} D'X_s \tag{4.80}$$

Proof

$$\begin{aligned} \Sigma &= \sigma^2 E(I + \lambda^{-1} N_s G N_s')E' \\ \Sigma^{-1} &= \frac{1}{\sigma^2} D'(I - N_s(N_s'N_s + \lambda G^{-1})^{-1} N_s')D \quad \text{by (2.19)} \\ &= \frac{1}{\sigma^2} D'(I - S + X_s(X_s'X_s)^{-1}X_s')D \quad \text{by (4.69).} \end{aligned}$$

(4.80) follows from (4.79) and (4.62). For y as in (4.77), the maximum likelihood estimates of β and β_s coincide with the estimates from the mixed model equations given in (4.70) and (4.71). ■

In the next lemma we give expressions for the matrices G, H, Q (as defined in section 3.2) for model (4.78).

Lemma 4.8 *The matrices G, H and Q for model (4.78) are*

$$G = \sigma^2 \begin{pmatrix} U & -UX'X_s(X_s'X_s)^{-1} \\ -(X_s'X_s)^{-1}X_s'XU & (X_s'X_s)^{-1} + (X_s'X_s)^{-1}X_s'XUX'X_s(X_s'X_s)^{-1} \end{pmatrix} \tag{4.81}$$

$$H = \frac{1}{\sigma^2}D'MD \tag{4.82}$$

$$Q = \frac{1}{\sigma^2}D'PD \tag{4.83}$$

where

$$U = (X'(I-S)X)^{-1} \tag{4.84}$$

$$M = X_s(X_s'X_s)^{-1}X_s' + (I-S)XUX'(I-S), \tag{4.85}$$

$$P = (I-S) + (I-S)XUX'(I-S)) \tag{4.86}$$

Proof From the definition of G given in (3.8),

$$\begin{aligned} G &= \left(\begin{bmatrix} X'E' \\ X_s'E' \end{bmatrix} \Sigma^{-1}[EX : EX_s] \right)^{-1} \\ &= \sigma^2 \begin{pmatrix} X'E'\Sigma^{-1}EX\sigma^2 & X'X_s \\ X_s'X & X_s'X_s \end{pmatrix}^{-1} \quad \text{by (4.80)} \\ &= \sigma^2 \begin{pmatrix} X'(I-S+X_s(X_s'X_s)^{-1}X_s')X & X'X_s \\ X_s'X & X_s'X_s \end{pmatrix}^{-1} \quad \text{by (4.79).} \end{aligned}$$

Expression (4.81) follows from the formula for the inverse of a partitioned matrix given in (4.13) and equations (4.79) and (4.80).

The definition of H given in (3.9) yields

$$
\begin{aligned}
H &= \Sigma^{-1}\left[\, EX \;:\; EX_s \,\right] G \begin{bmatrix} X'E' \\ X_s'E' \end{bmatrix} \Sigma^{-1} \\
&= \left[\, \Sigma^{-1}EX \;:\; \tfrac{1}{\sigma^2}D'X_s \,\right] G \begin{bmatrix} X'E'\Sigma^{-1} \\ \tfrac{1}{\sigma^2}X_s'D \end{bmatrix} \quad \text{by (4.80)} \\
&= D'\left[\, (I-S)XU \;:\; X_s(X_s'X_s)^{-1} - (I-S)XUX'X_s(X_s'X_s)^{-1} \,\right] \begin{bmatrix} X'E'\Sigma^{-1} \\ \tfrac{1}{\sigma^2}X_s'D \end{bmatrix} \\
&= \frac{1}{\sigma^2}D'[X_s(X_s'X_s)^{-1}X_s' + (I-S)XUX'(I-S)]D \\
&= \frac{1}{\sigma^2}D'MD \quad \text{from (4.85).}
\end{aligned}
$$

The proof of (4.83) is immediate from the definition of Q given in (3.10), (4.82) and (4.79). ■

Remark For Σ and G in (4.81) and (4.79), the estimates of β and β_s given by (3.12) coincide with the estimates given in (4.70) and (4.71) (obtained from the mixed model equations (4.67)).

In the next theorem, we apply the results given in section 3.5.3 to give the adjustments of the profile likelihood for model (4.78).

Theorem 4.4 *The bias and variance adjustments for the profile likelihood for model (4.78) are*

$$m'(\sigma^2,\lambda,\rho) = \left(-\frac{p+1}{2\sigma^2}, -\frac{1}{2\lambda}tr(M-V), tr(A)\right) \tag{4.87}$$

$$W_1(\sigma^2,\sigma^2) = W_2(\sigma^2,\sigma^2) = \frac{n-p-1}{2\sigma^4} \tag{4.88}$$

$$W_1(\sigma^2,\lambda) = W_1(\lambda,\sigma^2) = W_2(\sigma^2,\lambda) = W_2(\lambda,\sigma^2) = \frac{1}{2\sigma^2\lambda}tr(P-I+V) \tag{4.89}$$

$$W_1(\sigma^2,\rho) = W_1(\rho,\sigma^2) = W_2(\sigma^2,\rho) = W_2(\rho,\sigma^2) = -\frac{1}{\sigma^2}tr(B) \tag{4.90}$$

$$W_1(\lambda,\lambda) = W_2(\lambda,\lambda) = \frac{\lambda^{-2}}{2}tr\left((P-I+V)^2\right) \tag{4.91}$$

$$W_1(\lambda,\rho) = W_1(\rho,\lambda) = W_2(\lambda,\rho) = -\frac{1}{\lambda}tr((I-V)(I-P)EW) \tag{4.92}$$

$$W_2(\rho,\lambda) = -\frac{1}{\lambda}tr[(I-V-P)EW] \tag{4.93}$$

$$W_1(\rho,\rho) = tr(BB) - tr[(I+\lambda^{-1}N_sGN_s')E'W'C] + \frac{1}{\sigma^2}(X\beta+X_s\beta_s)'B'C(X\beta+X_s\beta_s) \tag{4.94}$$

$$W_2(\rho,\rho) = W_1(\rho,\rho) + tr(AB) \tag{4.95}$$

where

$$A = -VEW \tag{4.96}$$

$$B = -(I-V)WE \tag{4.97}$$

$$C = -PWE \tag{4.98}$$

and

$$V = X_s(X_s'X_s)^{-1}X_s' + (I - X_s(X_s'X_s)^{-1}X_s')XUX'(I-S). \tag{4.99}$$

Proof The proof follows from direct application of Lemmas 3.3, 3.4, 3.5, and 4.8 ■

Remark 1 The bias adjustment for σ^2 does not take into account the degrees of freedom (or number of effects, as it is called in the mixed model context) used to estimate the non-parametric part of the model. This is because the REML adjustment only accounts for the degrees of freedom corresponding to the fixed part of the model and in model (4.65), the smooth part of the model is consider as random.

4.5 Smoothing parameter selection

Most of the methods for selecting the smoothing parameter assume independent errors. Therefore, they generally tend to underestimate or overestimate the smoothing parameter, depending on the sign of the correlation parameter. Several authors (Altman, 1990; Hart, 1991) have proposed different methods to estimate the smoothing parameter in the presence of correlation when the smooth function is part of the mean. We present here two different approaches, one based of the *indirect method* of Altman (1990) and another based on the equivalence between the REML estimate of the smoothing parameter (when $f(z)$ is assumed to be random) and the modified GCV given by Wang (1998b).

4.5.1 Altman's indirect method

Altman (1990) proposed two methods for correcting the selection criterion when the correlation parameter is known. In our models, the competition parameter induces the correlation between the errors. In the context of section 4.1, i.e. the smooth term is fixed, the natural choice is the *indirect* method. The method uses a transformation of the residuals; in our case the transformed residuals for model (4.1) are

$$r = (I - M)Dy, \quad \text{for } M \text{ given in (4.12).}$$

Therefore the GCV criterion for the transformed residuals is

$$GCV = \frac{1}{n}\frac{y'D'(I-M)'(I-M)Dy}{(1-\mathrm{tr}(M)/n)^2} \tag{4.100}$$

This is equivalent to the ordinary GCV criterion for the adjusted response Dy. This method assumes the correlation parameter, ρ, to be known, and Altman (1990) uses the method of moments to calculate an estimate of ρ. In our case, the competition parameter induces the correlation but it is also present in the mean, and so we propose the use of the adjusted profile likelihood given in Theorem 4.1 to estimate it and (4.100) to estimate the smoothing parameter(s).

4.5.2 Wang's modified GCV

In the context of section 4.4, it is possible to use the adjusted profile likelihood to estimate λ and ρ. It is known (Silverman, 1985; Wang, 1998b) that the behavior of the estimates of λ obtained by GCV and REML are similar and Wang (1998b) gives the expression for the modified GCV which is the analog of the REML equations used to estimate λ. In the case of model (4.78) the modified GCV is given by

$$GCV = \frac{1}{n}\frac{y'D'(I-P)'D'D(I-P)Dy}{(\mathrm{tr}(D'(I-P)D/n)^2}, \tag{4.101}$$

for P given in (4.86).

Therefore, we will use an iterative procedure to estimate λ and ρ using the adjusted

profile likelihood or using (4.101) for λ and the adjusted profile for ρ.

Remark Wang (1998b) uses (4.101) to estimate λ and ρ simultaneously. This is only possible if ρ is the usual correlation parameter (i.e. only present in the error term). In the case of competition ρ is also present in the mean and the ordinary REML estimate is not unbiased, and therefore, it is necessary to use the adjusted profile likelihood to estimate ρ.

4.6 Standard errors of treatments in the presence of competition

In section 2.4 we presented an approximate method for the calculation of the standard errors of treatment estimates in a semiparametric additive model. Here we extend this method to semiparametric models in the presence of a competition effect.

Model (4.1): $y = X\beta + f(t) + \rho Wy + \epsilon$

The estimated treatment effects for model (4.1) are given in (4.16), and hence the variance is given by

$$\text{Var}(\hat{\beta}) = FDEE'D'F'\hat{\sigma}^2 = FF'\hat{\sigma}^2 \qquad (4.102)$$

where F is given in (4.18) as $F = (X'(I-S)X)^{-1}X'(I-S)$. Therefore the approximation given in (2.53) and (2.55) may be used.

Model (4.36): $y = JX\beta + f(z) + \epsilon$

This model is a special case of an ordinary semiparametric model, where the design matrix has been modified. The variance of $\hat{\beta}$ is given from (4.45) by

$$\text{Var}(\hat{\beta}) = F_J F_J'\hat{\sigma}^2 \qquad (4.103)$$

for F_J given in (4.47), $\hat{\sigma}^2 = y'(I-M_J)'(I-M_J)y/(n-p)$ and M_J given in (4.44). Again, expressions (2.53) and (2.55) may be used, substituting X by JX, for $J =$

$I + \rho W$.

Model (4.65): $y = X\beta + X_s\beta_s + N_s u_s + \rho W y + \epsilon$

The estimate of $\hat{\beta}$ for model (4.65) coincides with the estimate for model (4.1). However, in model (4.65) we use an stochastic approach to represent the smoother and this yields a different expression for the variance of $\hat{\beta}$. From (4.78)

$$\text{Var}(\hat{\beta}, \hat{\beta}_s) = \left([X : X_s]' \Sigma^{-1} [X : X_s]\right)^{-1},$$

which yields

$$\text{Var}(\hat{\beta}) = (X'(I - S)X)^{-1}\hat{\sigma}^2 \text{ by (4.81) and (4.84).} \quad (4.104)$$

The variance of $\hat{\beta}$ given in (4.102) and (4.104) would be equivalent if S was idempotent. Kenward and Roger (1997) gave adjustments for fixed effects in model (4.78); it would be of interest to see how the adjusted $se(\hat{\beta})$ compare with the ones obtained from (4.102).

Remark The calculation of the standard errors for models (4.1), (4.36) and (4.65) is easily extended to the case of q smoothers by substituting S, the smoother matrix, with M_q, the centred hat matrix of an additive model with q smooth terms.

4.7 Examples

In sections 3.5.2 and 3.5.3 we analysed the data given in Jenkyn et al. (1979). In both cases we accounted for the spatial variation using blocks and in both cases the competition parameter was significantly different from 0. We pointed out in section 3.5.3 (also indicated in Kempton, 1984) that competition and trend might be confounded in that data set. In this section we analyse the data using splines and locally weighted running lines to account for the trend. Two smoothers were used to analyse this data set, splines and loess. The data were tested for competition between neighbouring plots (sections 4.2 and 4.4) and competition between neighbouring treatments (section 4.3). Models (4.1) and (4.36) used both splines and loess, while model (4.65)

used smoothing splines only.

Table 4.1 gives estimated values and standard errors for ρ; the competition parameter is not significant in any case. An iterative procedure was used for the estimation of the smoothing and competition parameters in model (4.1) and (4.36), alternating between (4.100) or ordinary GCV for choosing the smoothing parameter, and Theorems 4.1 and 4.2 for the competition parameter. Model (4.65) deserves special attention: in this case we estimated the smoothing parameter and the competition parameter simultaneously.

The standard error of $\hat{\rho}$ is quite large (0.3) when model (4.65) is used. A possible explanation is that the competition and smoothing parameters are highly correlated. The covariance matrix for $\hat{\sigma}^2$, $\hat{\lambda}$ and $\hat{\rho}$ is

$$\mathrm{Cov}(\hat{\sigma}^2, \hat{\lambda}, \hat{\rho}) = \begin{pmatrix} 0.00004 & 0.00319 & 0.00143 \\ 0.00319 & 0.34865 & 0.15829 \\ 0.00143 & 0.158298 & 0.09375 \end{pmatrix} \tag{4.105}$$

and $\mathrm{Corr}(\hat{\lambda}, \hat{\rho}) = 0.853$. The positive value of the correlation coincides with what we would expect: when the smoothing parameter increases (fewer degrees of freedom, smoother trend), the competition parameter increases to account for the trend which has not been accounted for.

Smoother	Method	$\hat{\rho}$	se($\hat{\rho}$)	$\hat{\sigma}^2$	df
Spline	Model (4.1)	0.16	0.15	0.013	15.1
	Model (4.36)	0.04	0.03	0.009	15.4
	Model (4.65)	0.11	0.30	0.011	14.2
Loess	Model (4.1)	0.20	0.15	0.015	12.1
	Model (4.36)	0.04	0.03	0.010	14.9

Table 4.1: *Values of $\hat{\rho}$, se($\hat{\rho}$), $\hat{\sigma}^2$, and degrees of freedom for the smoother for Jenkyn data*

Remark The fact that the smoothing and competition parameter may be estimated simultaneously in model (4.65) does not ensure that both effects are identifiable. This agrees with Wang (1998b) who pointed out that this method does not perform well when the sample size is small, the trend varies quickly and/or the correlation parameter is large. Problems identifying correlation and trend are also commented on by Gampe (1998) *"the relation between the curvature of the true mean function, the strength of correlation and the variance error component determines the quality of the results"*.

Chapter 5

Analysis of cereal cultivar field trials

5.1 Introduction

In chapter 1 we gave a summary of the different approaches to the analysis of agricultural field experiments. Here we present the analysis of two barley trials and use the spatial model proposed by Wilkinson et al. (1983): *trend+error*. We use an additive model to represent the environmental trend (for example fertility in the soil in the case of yield or a gradient in disease), and use the results in chapter 4 to test for competition effects. In section 5.2 we present several approaches for the selection of the smooth component of a semiparametric additive model, and we illustrate them with the analysis of a winter barley trial which had trend induced deliberately. In section 5.3 we analyse a large spring barley trial with a more conventional design. In this trial, we focus on the assessment of the competition effect and address a fundamental problem in the analysis of agricultural trials: which model is the appropriate one?

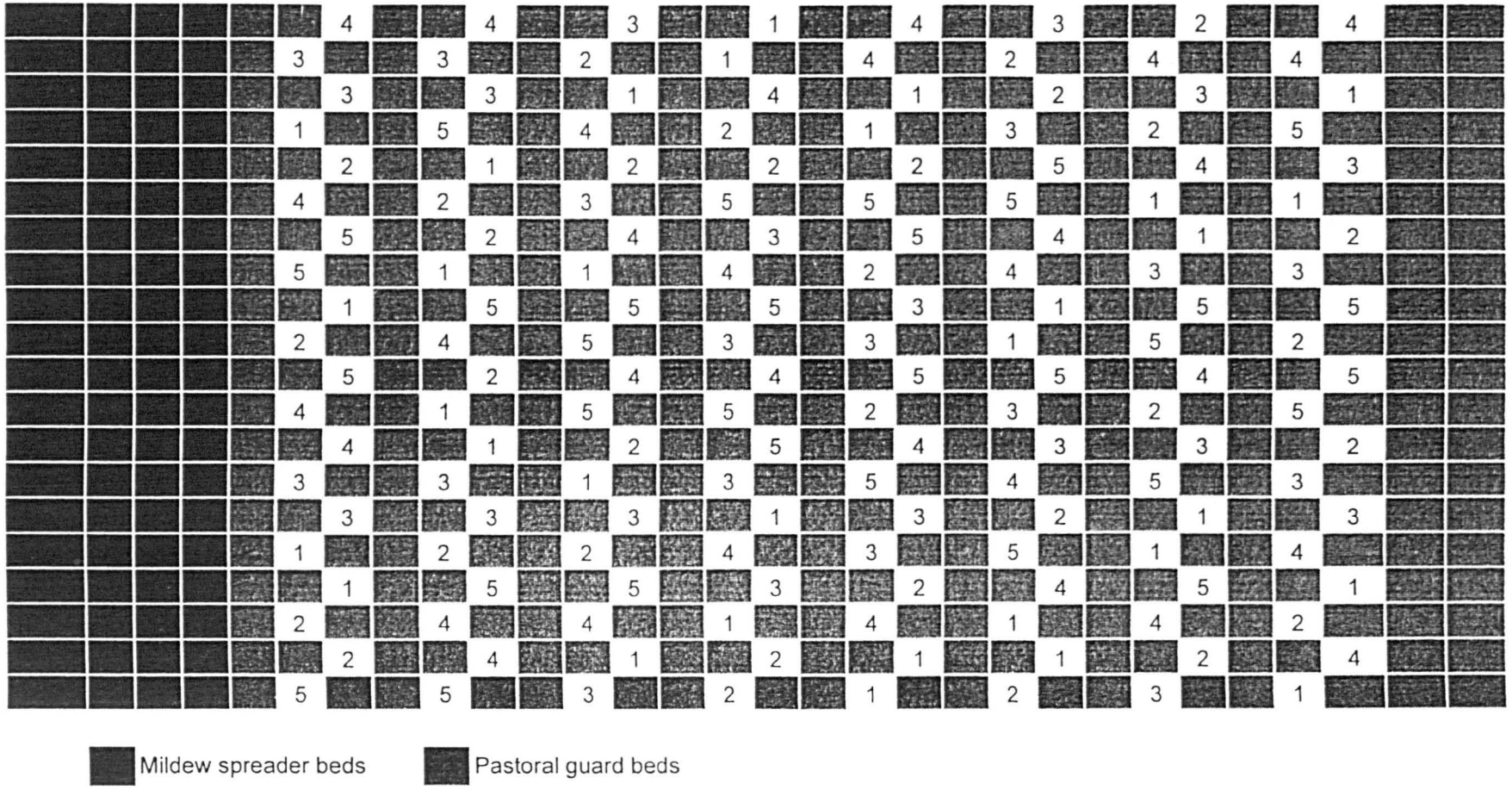

Figure 5.1: *Field plan for BB trial.*

5.2 Strategies for selecting the spatial component of a semiparametric additive model

In this section we present several methods for selecting the smooth term in a semiparametric model and apply them to the analysis of an agricultural trial.

5.2.1 Layout of trial BB

A winter barley trial (trial BB) was grown at the Scottish Crop Research Institute, Dundee in September 1995. Figure 5.1 gives the trial layout, with the left side of the diagram corresponding to the west. The trial consisted of 16 entry columns arranged in a split-plot design, and 14 guard or spreader beds. Each column was 1.55 m wide and 50 m long. The 16 entry columns were grouped into four blocks each of four columns. Each column was divided into 20 plots of length 2.5 m, and after plot demarcation and allowing for wheel tracks, the final plot size was 1.22 m $\times$ 1.95 m. Two replicates of five test cultivars (Pipkin, Puffin, Fighter, Manitou and an equal proportion mixture of these) were grown on alternate plots in each block, while the cultivar Pastoral was grown on the remaining plots giving a chequer-board design. The 14 guard or spreader columns were arranged in the following way: guard beds of the cultivar Pastoral were grown between each main plot as protection against fungicide drift, and two further guard beds of Pastoral were grown on the east side of the trial. Finally, to encourage a trend in the level of powdery mildew, four columns of the susceptible spring barley Golden Promise were grown on the west (prevailing wind) side of the trial. A foliar fungicide was applied to one main plot in each block when powdery mildew first started to appear on Golden Promise. This gave almost complete control of the disease in all genotypes in the fungicide-treated columns for the whole season.

Spatial variation in the yield and mildew severity was encouraged in two ways. First, the trial was grown on an exposed site and so the four columns of Golden Promise on the windward side of the trial promoted a gradient of powdery mildew from west

to east. Second, a fertility gradient was introduced by dressing plots in the southern half of the trial with steadily increasing amounts of fertiliser up to twice the normal rate for such trials.

All plots, except the complete guard and spreader beds but including check plots, were harvested. A small plot combine was used and the grain yield recorded at 9% moisture content. Five measurements of mildew level were also recorded during the season, but we will concentrate on the analysis of yield.

5.2.2 Semiparametric modelling

In the context of field trials, semiparametric additive models use one or more smoothers to account for the spatial variation in, for example, yield as functions of the row and column position of the plots in the field. Other crop measurements, such as height or disease severity, may also be used in the model.

Two analyses were carried out for this data set: analysis of the full trial and analysis of all plots except the check plots. In the next sections we give several strategies on how to choose between these two models

$$y_{ijk} = \mu + \tau_j + \alpha_k + (\alpha\tau)_{jk} + lo(r,b) + \epsilon_{ijk}. \tag{5.1}$$

$$y_{ijk} = \mu + \tau_j + \alpha_k + (\alpha\tau)_{jk} + lo(r) + lo(b) + \epsilon_{ijk} \tag{5.2}$$

where y_{ijk} is the yield of the plot i in row r and column b, which is planted with cultivar j, $j = 1,\ldots,6$, and receiving fungicide ($k = 1$) or no fungicide ($k = 2$), $lo(r,b)$ and $lo(r) + lo(b)$ represent a two-dimensional loess smoother or the sum of two one-dimensional loess smoothers. The amount of smoothing and the dimension of the smoother will affect the estimated smooth function and, therefore, the overall fit of the model. In the next section we present two simple graphical tools to check whether a trend is present and whether it is additive or two-dimensional.

5.2.3 Graphical methods

A useful first check for the presence of trend in agricultural trials is to calculate the residuals from a model with all treatment effects but no spatial component, and plot them against row and column position. Figure 5.2 shows plots of these partial residuals against row and column position. A scatterplot smoother was added to indicate the possible trends. There is an obvious trend in the plot against row position with the residuals increasing as the row number increases. The plot against column position shows an increase of the residuals in the left half of the trial and smaller residuals on the right-hand side. These two plots suggest that a trend down the rows and across columns may be present. However, this assumes that the trend affecting this trial may be decomposed into the sum of two trends. In field trials this is not always the case, and further investigation is necessary to check whether the trend is better expressed as a two-dimensional surface. We use conditional plots or **coplots** (Cleveland, 1993) as a visual tool for studying how a response depends on two or more covariates. In the context of agricultural trials, we use them to detect whether the trend across the columns varies with the row number (or vice-versa), suggesting that

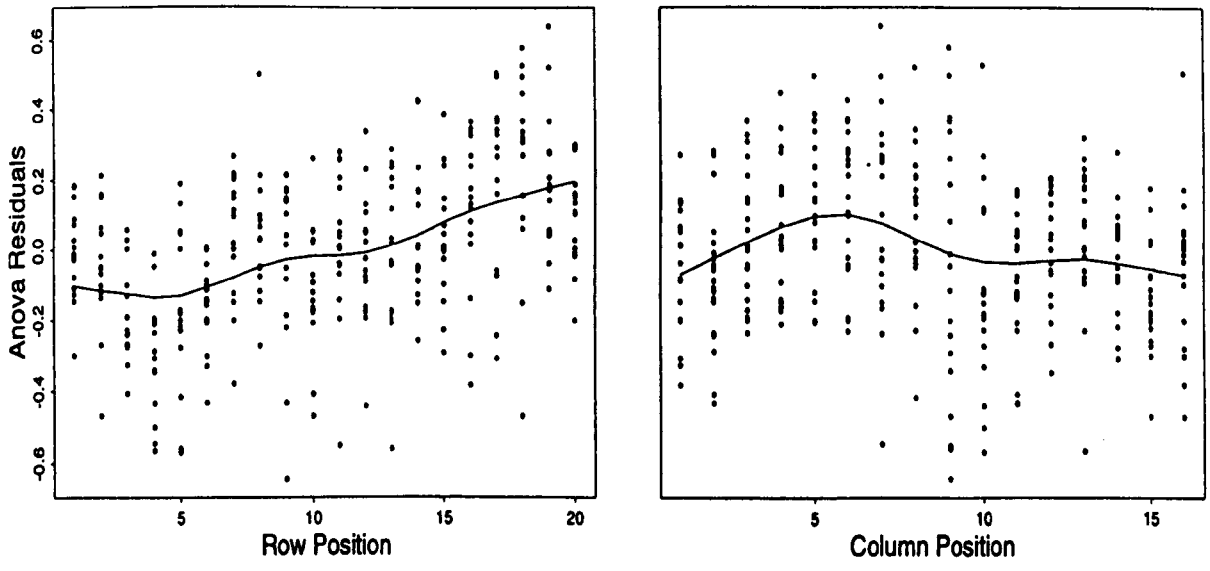

Figure 5.2: *Plot of residuals against row and column position with scatterplot smoother.*

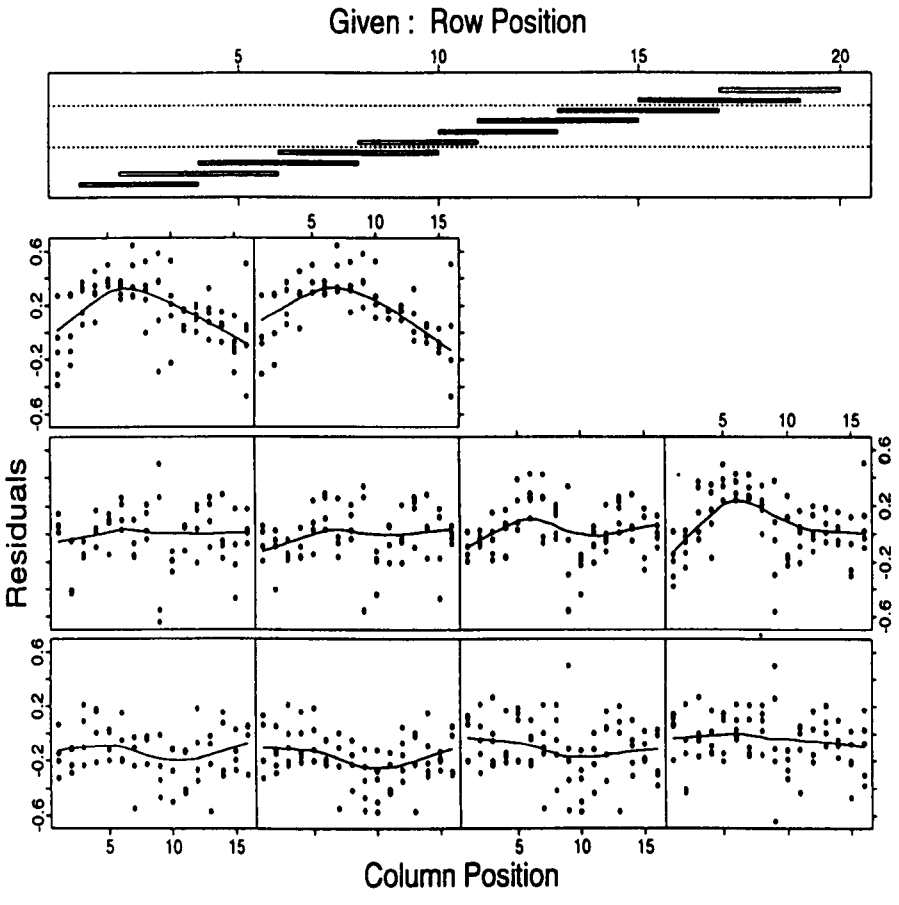

Figure 5.3: *Conditional plots of partial residuals versus column for different row positions.*

a two-dimensional trend is present. We plot the partial residuals against, for example, column number, for values of row number in a certain interval. The panels at the top part of the plot show these intervals. There are the same number of points in each interval. If the relationship between the column position and the partial residuals changes with the row number, then this will indicate that there is an interaction between row and column position and therefore, a two-dimensional smoother might be more appropriate. Figure 5.3 shows no evidence of a trend in the lower half of the trial, but there is a strong increase and then a decrease of the residuals for plots in the top rows of the trial. Residuals were larger in the top part of the trial compared with the residuals for plots in rows 1-10. This agrees with the trend expected from the fertility gradient that was introduced in this trial. We may conclude from here that

a two-dimensional smooth surface might be required to describe the spatial variation in this trial.

5.2.4 Analytical Methods

If the partial residual plots indicate that one or more smooth terms should be included in the model, it will be necessary to select the correct degree of smoothing. For a *loess* smoother, the amount of smoothing is given in terms of the span of the smoother. The span may be chosen by minimisation of some appropriate criterion. We present four of them: cross-validation (CV) (Stone, 1974), generalised cross-validation (GCV) (Craven and Wahba, 1979), the Akaike information criterion (AIC) (Akaike, 1973), and recently, a modified version of the Akaike criterion (AIC_C) (Hurvich et al., 1998; Simonoff and Tsai, 1999).

Cross-Validation

Let $\hat{\beta}^{-i}$ and $\hat{f}_j^{-i}$, $j = 1, \ldots, q$ be the estimates of β and f_j in the semiparametric additive model $y = X\beta + f_1 + \ldots + f_q + \epsilon$ with the $i\text{-}th$ data point omitted. The fitted value of y_i at the $i\text{-}th$ data point based on the remaining data is then

$$\hat{y}_i^{(-i)} = X_i'\hat{\beta}^{(-i)} + \sum_j \hat{f}_j^{(-i)}(z_{ij}).$$

We define $\lambda = (\lambda_1, \ldots, \lambda_q)'$ where λ_j is the smoothing parameter of f_j. The cross-validation score is defined by

$$CV(\lambda) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i^{(-i)})^2.$$

Efficient calculation of $CV(\lambda)$ is possible since

$$y_i - \hat{y}_i^{(-i)} = \frac{y_i - \hat{y}_i}{1 - H_{ii}}$$

and thus,

$$\mathrm{CV} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{y}_i}{1 - H_{ii}}\right)^2 \tag{5.3}$$

where H_{ii} are the diagonal entries of the hat-matrix ($H = (h_{ij})$) for model (2.40) given in (2.45). The cross-validation criterion, CV, is computationally very demanding if

the data set is large and/or there are several smooth terms. We present other criteria based on different functions of the trace of H which are much simpler to compute.

Generalised Cross-Validation

Generalised cross-validation GCV is based in the same principle as cross-validation, but each element H_{ii} is replaced with its average value, $\mathrm{tr}(H)/n$:

$$\mathrm{GCV} = \frac{1}{n} \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{[1 - \mathrm{tr}(H)/n]^2}. \tag{5.4}$$

This is a generalisation of the expression for the GCV criterion for semiparametric models with one smooth term given by Green et al. (1985).

Akaike Information criterion

In the Akaike Information criterion, the smoothing parameter is chosen to minimise

$$\mathrm{AIC} = \log \hat{\sigma}^2 + 2\mathrm{tr}(H)/n \tag{5.5}$$

where

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{j=1}^{n} (y_i - \hat{y}_i)^2.$$

AIC may be used to compare models which are not nested. However, it is well known (see for example Hurvich and Tsai, 1989) that AIC under-smooths the data in many cases.

Modified AIC

Hurvich and Tsai (1989) proposed a modified version of (5.5) to avoid under-smoothing of the data. This was extended by Hurvich et al. (1998) to the case of non-parametric regression and Simonoff and Tsai (1999) applied it to semiparametric and additive model selection

$$\mathrm{AIC}_C = \log \hat{\sigma}^2 + 1 + \frac{2(\mathrm{tr}(H) + 1)}{n - \mathrm{tr}(H) - 2}. \tag{5.6}$$

In many cases a decrease in the span corresponds to a small change in the criteria. We present two methods for checking the significance of those changes.

Approximate F test

Several authors (Cleveland and Devlin, 1988; Hastie and Tibshirani, 1987, 1990) have discussed inference methods in analogy with linear regression. Suppose we want to compare two semi-parametric models with spans λ_1 and λ_2 ($\lambda_2 < \lambda_1$). (If the model has more than one smooth term, then λ_1 and λ_2 are vectors and $\lambda_2 \leq \lambda_1$ with strict inequality for at least one component). Let H_1 and H_2 be the hat matrices associated with those two semiparametric models with spans λ_1 and λ_2. We seek a test of the null hypothesis: the span is λ_1 against the alternative hypothesis: the span is λ_2. Let $RSS_i = y'(I - H_i)'(I - H_i)y$, $i = 1, 2$ be the residual sum of squares for the two fits and $\gamma_1 = \mathrm{tr}(R_1)$ and $\gamma_2 = \mathrm{tr}(R_2)$, for $R_i = (I - H_i)'(I - H_i)$. Then a first approximation (Hastie and Tibshirani, 1990) uses

$$F_1 = \frac{(RSS_1 - RSS_2)/(\gamma_2 - \gamma_1)}{RSS_2/\gamma_2} \sim F_{\gamma_2 - \gamma_1, \gamma_2}. \tag{5.7}$$

In practice, we will use the approximation for the degrees of freedom associated with the hat matrix of an additive model suggested by Buja et al. (1989) and used in most statistical packages, and define

$$\gamma_i \sim \mathrm{tr}(I - H_i). \tag{5.8}$$

(5.7) may be improved by a two-moment correction (Cleveland and Devlin, 1988) which yields

$$F_1 \sim F_{v_1^2/v_2, \delta_1^2/\delta_2} \tag{5.9}$$

where $v_1 = \mathrm{tr}(R_1 - R_2)$, $v_2 = \mathrm{tr}(R_2 - R_1)^2$, $\delta_1 = \mathrm{tr}(R_2)$ and $\delta_2 = \mathrm{tr}(R_2^2)$. Hastie and Tibshirani (1990) reported the degrees of freedom of the numerator are changed significantly by this adjustment. Our experience with several examples agrees with this, suggesting that (5.9) is more appropriate than (5.7). However, the manipulation of the $n \times n$ matrices H_i required to compute v_2 and δ_2 is computationally very demanding.

Bootstrap F test

As an alternative, we propose a test using the original F_1 (with γ_i given in 5.8), but bootstrap the residuals from the model under the null hypothesis to obtain its

Span	CV	GCV	AIC	AIC_C
6/320	.0473	.0484	-3.440	-1.323
16/320	.0329	.0329	-3.487	-2.309
26/320	.0312	.0313	-3.501	-2.414
36/320	.0314	.0314	-3.479	-2.428
46/320	.0320	.0320	-3.460	-2.419
56/320	.0332	.0331	-3.419	-2.388
66/320	.0338	.0337	-3.399	-2.372
76/320	.0345	.0344	-3.378	-2.353

Table 5.1: *Comparison of criteria for selection of the span, in a two-dimensional semi-parametric additive model for trial BB.*

distribution. Hastie and Tibshirani (1990) also estimated exact tail probabilities by applying this method to binary data. One of the main advantages of this method is that it can estimate the distribution of the test statistic for non-nested hypotheses, that is, cases in which the null model is not a sub-model of the alternative. In these cases, the quadratic forms in the numerator and denominator are not independent and standard results cannot be applied.

In the context of field trials, we will use the bootstrap F test to choose the best additive trend, the best two-dimensional trend and to choose between them.

5.2.5 Analysis of trial BB

Figure 5.2 showed that there was a trend down the rows and across columns and Figure 5.3 suggested that the trend might be two-dimensional. CV, GCV, AIC and AIC_C were calculated for a range of spans for a two-dimensional smoother (Table 5.1). The minimum value of the AIC_C was obtained at a span of 36/320, the minimum of the other three criteria was obtained at a span of 26/320. The approximate F test given in (5.7) with γ_i as in (5.8) was used to compare models with spans 26/320 and 36/320 respectively. The test statistic had a value of 2.18 compared with the 95

and 99% points of 1.76 and 2.20 for an F distribution with 13.5 and 263.3 degrees of freedom. The 95% point of the bootstrap distribution was 2.18 and the corresponding 99% point was 2.40. Hence, the evidence for an improvement in fit associated with a decrease in the span from 36/320 to 26/320 is borderline and the approximate F test would provide stronger evidence for the smaller span than the data warrants. A further bootstrapped F test was used to compare models with span 46/320 and 36/320, the value was 2.81 compared with the 95 and 99% points of 2.35 and 2.59. We conclude that a two-dimensional loess with span 36/320 is appropriate for this data set.

A model with two one-dimensional smooth terms (for rows and columns) was also fitted. To reduce the computational cost CV was not used, but GCV, AIC and AIC_C indicated that the spans for the best additive model were 5/20 (for rows) and 7/16 (for columns). Bootstrapped F tests confirmed that there was not any significant improvement of the fit by increasing or decreasing the spans. A further bootstrap F test was used to compare this model with a model with a two-dimensional loess (span=36/320). The value of the test statistic was 4.40 and the 95 and 99% points of the bootstrapped distribution were 1.67 and 2.03 respectively, indicating that a model with a two-dimensional smoother is necessary (which confirms the conclusion drawn from the coplot). Figure 5.4 shows the fitted surface for both additive and

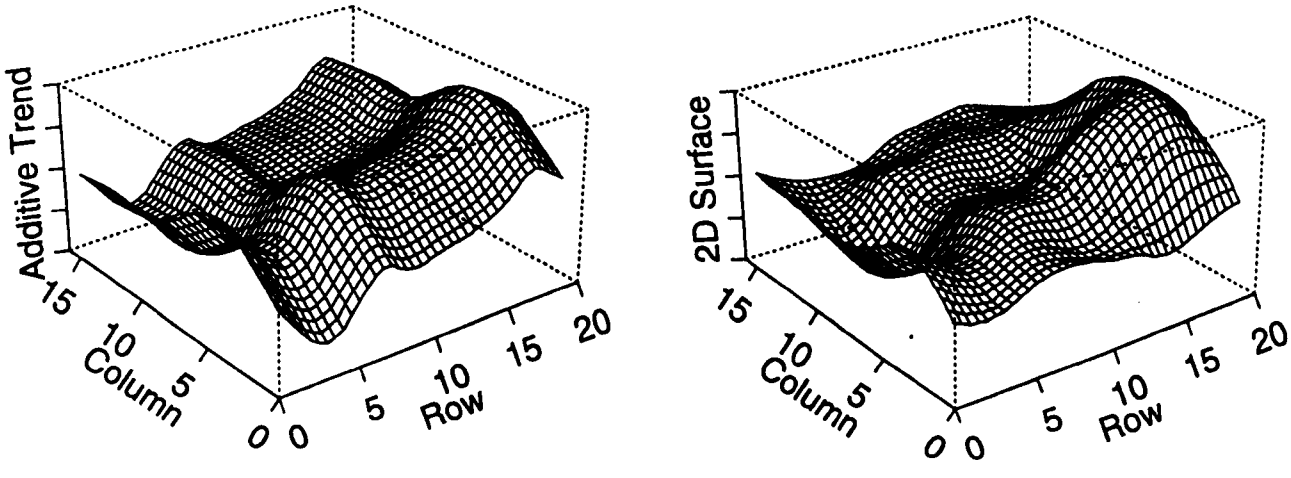

Figure 5.4: *Plot of additive and two-dimensional trend for trial BB.*

Model	*Res S.S.*	*Res d.f.*	*Test*	*Change in d.f.*	*v.r.*	*F Pr.*
(a) lo(row,col)+C+F+C.F	7.527	276.8	-	-	-	-
(b) lo(row,col)+C+F	7.712	281.8	(a) *v* (b)	5	1.36	n.s
(c) lo(row,col)+C	7.737	282.8	(b) *v* (c)	1	0.92	n.s
(d) lo(row,col)+F	12.227	286.8	(b) *v* (d)	5	33.21	<0.001
(e) lo(row,col)	12.265	287.8	(d) *v* (e)	1	1.40	n.s
			(c) *v* (e)	5	33.30	<0.001

Table 5.2: *Anova table for the significance of the treatment effects for BB trial with all plots. C= cultivar, F= fungicide effect. The two-dimensional smooth term lo(row,col) has span 36/320.*

two-dimensional trends. The high yields in the high row positions correspond to the area of high fertiliser application.

The significance of the treatments was investigated by dropping treatments terms to build up an analysis of variance table. A bootstrap F test was also used in this case. Due to the non-parametric nature of the smooth terms, the treatments are no longer orthogonal, so different orders for the removal of the terms should be compared. Table 5.2 summarises these comparisons. Yield was not significantly affected by the presence or absence of fungicide and the change in residual sum of squares due to the exclusion of cultivar or treatment effect was affected very little by the other terms in the model.

From the experimental point of view, our main interest is in the estimation and standard errors of the treatment effects and our objective is to improve the precision of the estimates by using semiparametric additive models. The variance of the treatment effects was calculated using the method shown in section 2.4.2. The average standard error of difference for comparisons of two cultivars (excluding the check plot cultivar) was 0.042, compared with 0.023 for fungicide effect (see Table 5.3 for culti-

Treatment		*Additive model*		*ANOVA*		*ANCOVA*
		Check plots	*No check plots*	*Check plots*	*No check plots*	*Check plots*
Cultivar	Pipkin	1.451	1.453	1.463	1.463	1.460
	Puffin	1.122	1.126	1.114	1.114	1.071
	Fighter	1.284	1.276	1.287	1.287	1.281
	Manitou	1.568	1.574	1.575	1.575	1.565
	Mixture	1.475	1.473	1.465	1.465	1.478
	Pastoral	1.490	-	1.490	-	-
S.E.D.		0.042	0.046	0.050	0.052	0.044
Disease control	Fungicide	1.450	1.421	1.470	1.427	1.390
	No fungicide	1.417	1.339	1.401	1.334	1.362
S.E.D		0.023	0.032	0.050	0.046	0.028

Table 5.3: *Treatment means and standard errors for each model.*

var and fungicide means). Figure 5.5 shows that the estimates of cultivar effects are robust with respect to the change in the span. A conditional plot (Figure 5.6) of the residuals after fitting the semiparametric model with the two-dimensional loess smoother showed that there was no trend left unexplained.

The analysis was repeated, omitting the check plots. A bootstrap F test showed that a two-dimensional smoother with span 28/160 gave a better fit to the data than two one-dimensional smooth terms. The GCV criterion yielded the same conclusion, in comparison with AIC_C which had a minimum at span 38/160. Table 5.4 indicates the significance of the treatment effects. In this case, the fungicide had a significant effect on the yield. (See Remark 1 on page 112 for a comment on this finding.) Cultivar means (Table 5.3) are similar to the means from the analysis of the full data set. The average standard errors for cultivar and fungicide effect increased slightly (0.046 and 0.032 respectively) compared with the standard errors from the full analysis. Figure 5.7 shows that the fitted surface is very similar to the one obtained using data on all plots (Figure 5.4).

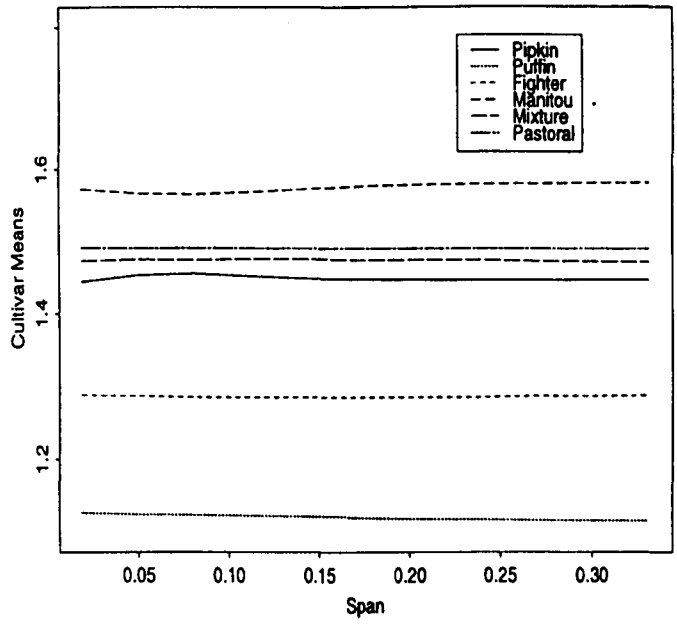

Figure 5.5: *Plot of cultivar means for different spans.*

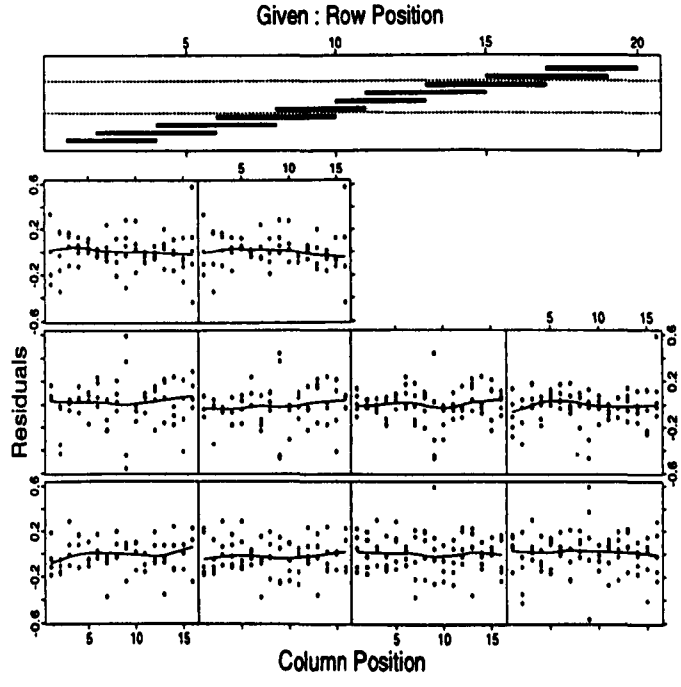

Figure 5.6: *Conditional plots residuals, after fitting model (5.1) versus column position for different row positions.*

Model	*Res S.S.*	*Res d.f.*	*Test*	*Change in d.f.*	*v.r.*	*F Pr.*
(a) lo(row,col)+C+F+C.F	4.137	128.4	-	-	-	-
(b) lo(row,col)+C+F	4.281	132.4	(a) *v* (b)	4	1.12	n.s
(c) lo(row,col)+C	4.390	133.4	(b) *v* (c)	1	3.38	<0.05
(d) lo(row,col)+F	7.835	136.4	(b) *v* (d)	4	27.58	<0.001
(e) lo(row,col)	7.988	137.4	(d) *v* (e)	1	4.75	<0.05
			(c) *v* (e)	4	27.92	<0.001

Table 5.4: *Anova table for the significance of treatment effects of BB trial without the check-plots. The two-dimensional smooth term has span 38/160.*

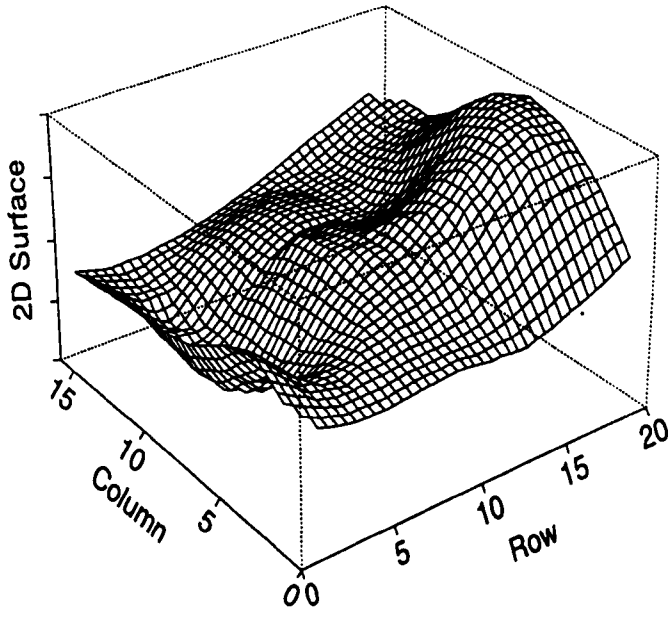

Figure 5.7: *Plot of two-dimensional loess with span 28/160 for trial BB without the check plots.*

5.2.6 Comparison with other spatial analyses

We compare the semiparametric analysis with other classic methods of analysis for spatial data: analysis of variance, taking into account the split plot design and a

check plot analysis using the mean of the neighbouring plots as a covariate to represent fertility.

Analysis of Variance

Table 5.5 shows the mean squares and variance ratios and the significance of the cultivar and fungicide effects are similar to the semiparametric analysis in both cases

Source of variation	*d.f.*	*m.s.*	*v.r.*	*F Pr.*
Block	3	0.481		
Fungicide	1	0.377	1.86	0.226
Residual	3	0.203		
Sub-plot	24	0.16		
Cultivar	5	1.032	25.59	<0.001
Interaction	5	0.016	0.41	0.839
Residual	278	0.040		
Block	3	0.124		
Fungicide	1	0.351	4.13	0.135
Residual	3	0.085		
Sub-plot	24	0.126		
Cultivar	4	1.051	24.50	<0.001
Interaction	4	0.008	0.18	0.946
Residual	120	0.043		

Table 5.5: *Analysis of variance of the yields of trial BB, with (top) and without (bottom) check plots.*

(with and without check plots). The differing sizes of block, whole plot and sub-plot residual mean squares indicate that this model leaves some spatial variation unexplained. The standard errors of cultivar effect reported in Table 5.3 are larger than the ones in the corresponding semiparametric analysis.

Check plot analysis

The check plot analysis uses the yield from neighbouring check plots to adjust for a possible gradient in fertility. In this trial, a standard variety, Pastoral, was chosen for the check plots. For each plot, the covariate was calculated by averaging of the neighbouring check plots receiving the same fungicide treatment. The variance components corresponding to block and whole plots were not significantly different from zero, therefore an ordinary analysis of covariance model was used,

$$y_{ijk} = \mu + \tau_j + \alpha_k + (\alpha\tau)_{jk} + \theta(Z_{ijk} - \bar{Z}) + \epsilon_{ijk}. \tag{5.10}$$

Significance and means of treatment effects (see Table 5.3) were similar to the ones obtained in the analysis of variance.

Remark 1 All analyses yield similar conclusions about the cultivar treatment. However, fungicide effect was not significant when all plots (including check plots) were used in the analysis of variance and semiparametric model, and it was significant when the check-plot analysis and semiparametric model was used to analyse the data without the check plots. A possible explanation for this is the fact that the check plots account for half of the replication, and the means with and without fungicide for the check plots, 1.51 and 1.47, are not significantly different and this is dominating the fungicide effect in the analysis of the full trial. Moreover, we believe it is not appropriate to compare the significance of the fungicide treatment in the split-plot analysis with the semiparametric model, as the F distributions used in each case have very different degrees of freedom. The correct test for a main-plot treatment, such as fungicide, is not obvious when a semiparametric model is applied to a split-plot design. We will discuss this in more detail in section 5.3 and chapter 6.

Remark 2 In all cases, the semiparametric additive model gave smaller standard errors: when the information in all plots was used, the semiparametric model performed better that the split-plot analysis and the analysis of covariance; when the check plots were ignored, the semiparametric model also gave more accurate estimates of the standard errors of treatments than the split-plot analysis. In some previous trials at the Scottish Crop Research Institute, check-plots have been used to either raise or lower the overall level of disease in a trial (according to whether a susceptible or resistant variety is used) but their mildew or yields levels were not of direct interest, and so have not been recorded. By recording and analysing the yield of trial BB, we have shown that the semiparametric model applied to the data without the check plots gave a very similar fitted surface to the same model for the full data set. Further data analysis is needed, but our findings from these and other data analysed, indicate that if a semiparametric additive model is used to analyse the data, we do not need the mildew and yield scores in the check plots.

Remark 3 The cultivar means were quite robust with respect to the smoothing parameter and very similar to the means from the split-plot analysis. However, the estimates of treatment means might be altered more in an experiment with less replication.

5.3 Competition, trend or correlation?: Analysis of trial BJ

A spring barley trial was grown on the Gourdie farm of the Scottish Crop Research Institute in 1996. This is a more conventional trial and no trends have been induced. The design of this trial allowed us to study in more detail the comparisons with the analysis of variance and point out the effect that local correlation may have in a semiparametric model.

5.3.1 Layout of trial BJ

The trial consisted of 72 columns arranged in a split-split-split-plot design. The 72 columns were grouped into 3 blocks each of 24 columns. Each block was divided into two main plots of 12 columns and a foliar fungicide treatment (presence/absence) applied to each main plot. Each main plot was divided into two sub-plots of 6 columns and a nitrogen treatment with two levels was applied to each sub-plot. Sub-plots were divided into thirty units in a 10×3 layout, so that each unit was 2 columns wide by 1/10 column long. Thirty barley cultivars or cultivar mixtures were randomised within each sub-plot. The reason for having plots two columns wide was to obtain units more nearly square than usual, and so to investigate whether competition effects might be detected along or across columns (in trials with long, narrow plots, competition is expected only in one direction). The two columns comprising each unit were harvested separately, so we had information on spatial variability of the yield within each unit. The layout of one block is shown in Figure 5.8.

BLOCK 1

19	19	4	4	13	13	15	15	14	14	26	26	12	12	10	10	18	18	20	20	8	8	5	5
26	26	30	30	1	1	30	30	23	23	12	12	9	9	25	25	14	14	3	3	14	14	13	13
17	17	24	24	16	16	18	18	2	2	27	27	8	8	5	5	1	1	16	16	25	25	10	10
2	2	9	9	20	20	22	22	7	7	24	24	29	29	22	22	3	3	30	30	1	1	18	18
14	14	11	11	10	10	21	21	6	6	8	8	26	26	28	28	17	17	23	23	22	22	12	12
21	21	12	12	6	6	25	25	24	24	3	3	27	27	15	15	4	4	7	7	6	6	26	26
7	7	27	27	25	25	4	4	11	11	10	10	7	7	20	20	19	19	11	11	29	29	21	21
23	23	22	22	8	8	16	16	28	28	1	1	6	6	13	13	23	23	4	4	15	15	27	27
5	5	3	3	18	18	17	17	9	9	5	5	2	2	11	11	24	24	17	17	9	9	2	2
29	29	28	28	15	15	29	29	19	19	13	13	16	16	30	30	21	21	19	19	28	28	24	24

F1	F0

N1	N2	N2	N1

Figure 5.8: *Field plan of Block 1 for BJ trial.*

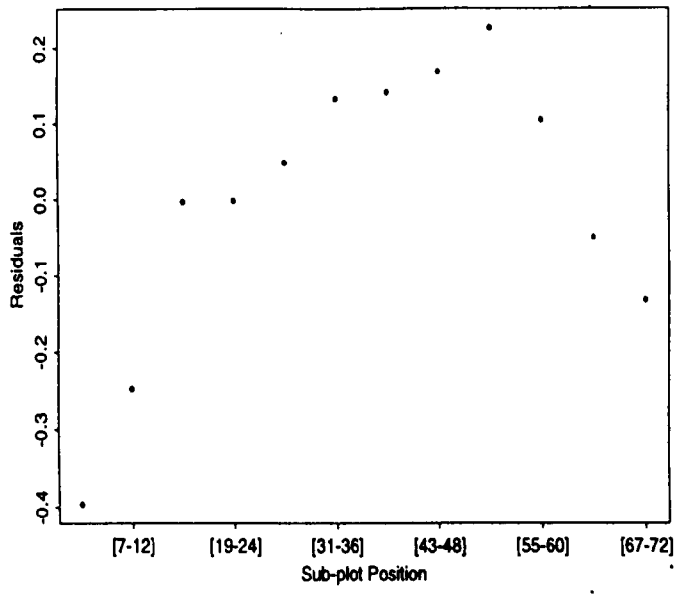

Figure 5.9: *Plot of partial residuals after analysis of variance of the sub-plot mean yields against sub-plot position. The numbers in brackets are the column numbers in each subplot.*

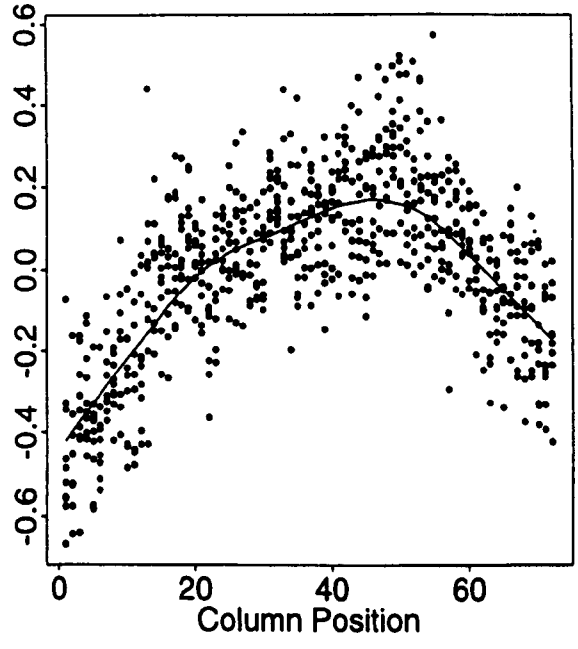

Figure 5.10: *Plot of residuals against column position with scatterplot smoother.*

5.3.2 Analysis of variance

A split-split-split-plot analysis (see Table 5.6) was used to analyse the yield data. Cultivar and nitrogen treatments and the interaction between cultivar and fungicide were significant. The residual mean square associated with the sub-plot was small (0.388) compared with that of the the main plot (5.156), and large compared with the unit stratum (0.037) indicating that a possible large scale trend may be present. To investigate the presence of the trend, a model with terms for fungicide, nitrogen and their interaction was fitted to the sub-plots means. Figure 5.9 is a plot of the partial residuals from that model against the column number in each sub-plot. There is a strong trend from the left to right side of the trial. This was confirmed by a plot of partial residuals against column position, after fitting all three factors (cultivar,

Stratum	*Source of variation*	*d.f.*	*m.s.*	*v.r.*	*F Pr.*
Block	Block	2	5.148		
Main-plot	Fungicide	1	35.760	6.94	0.119
	Residual	2	5.156		
Sub-plot	Nitrogen	1	9.166	23.60	0.008
	Fung:Nit	1	0.207	0.53	0.506
	Residual	4	0.388		
Unit	Cultivar	29	0.937	24.98	<0.001
	Fung:Cult	29	0.087	2.33	< 0.001
	Nit:Cult	29	0.041	1.08	0.356
	Fung:Nit:Cult	29	0.043	1.14	0.292
	Residual	232	0.037		
Half-unit	Half-unit	360	0.009		

Table 5.6: *Analysis of variance of the half-unit yields of trial BJ.*

nitrogen and fungicide) and their interaction to the half-unit yields (Figure 5.10). In the next section, we will fit a semiparametric model and estimate the trend with a nonparametric smoother.

5.3.3 Semiparametric modelling

The cultivar comparisons are at the unit stratum and we will fit our first semiparametric model using the yield on each unit as a response (i.e. summing the two half-unit yields). Later we will compare this with an analysis based on the individual half-units.

Analysis of unit yields

Partial residuals, after fitting a model with all three main effects and interactions to the units yields, were used in the conditional plot given in Figure 5.11. The plot

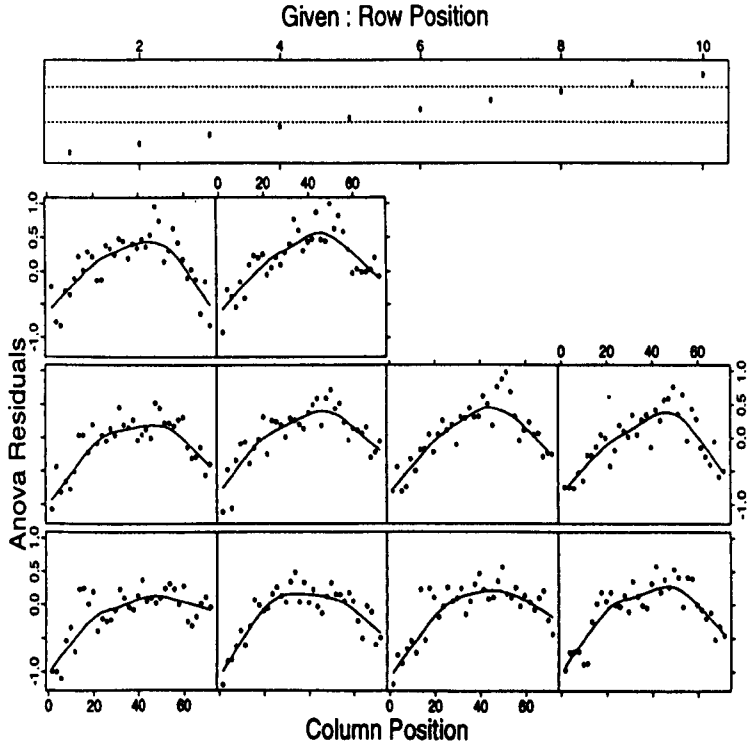

Figure 5.11: *Conditional plots of partial residuals versus column for different row positions for trial BJ.*

suggested the sum of two one-dimensional terms may suffice to explain the trend. A

model with two additive loess terms for rows (span 5/10) and columns (span 10/36) was fitted to the data. Figure 5.12 shows the estimated trend in both directions. A model with a two-dimensional loess smoother (span 38/360) was also fitted (see Figure 5.13). Both the approximate and bootstrapped F tests were used to compare the models, and both tests indicated that a two-dimensional smoother is necessary. Due to the complexity of the model, we give only one possible analysis of variance table (Table 5.7); there are many other ways to drop the different terms, but the significance of the treatments and interactions was not affected by the order in which the terms were dropped. All three main effects and the interaction between cultivar and fungicide were significant. The average standard error for differences between cultivars was 0.094, improving the one obtained in the corresponding analysis of variance

Model	*Res S.S.*	*Res d.f.*	*Test*	*Change in d.f.*	*v.r.*	*F Pr.*
(a) lo(row,col)+F+N+C+C.F+C.N+F.N+C.F.N	10.536	210.4	-	-	-	-
(b) lo(row,col)+F+N+C+C.F+C.N+F.N	12.293	239.4	(a) *v* (b)	29	1.28	n.s
(c) lo(row,col)+F+N+C+C.F+C.N	12.336	240.4	(b) *v* (c)	1	0.76	n.s
(d) lo(row,col)+F+N+C+C.F	13.791	269.4	(c) *v* (d)	29	0.98	n.s
(e) lo(row,col)+F+N+C	17.943	298.4	(d) *v* (e)	29	2.79	< 0.001
(f) lo(row,col)+F+N	63.321	327.4	(e) *v* (f)	29	26.02	< 0.001
(g) lo(row,col)+F	79.758	328.4	(f) *v* (g)	1	84.99	< 0.001
(h) lo(row,col)	99.913	329.4	(g) *v* (h)	1	82.99	< 0.01

Table 5.7: *Anova table for the significance of the treatment effects for the unit yields of BJ trial. The two-dimensional smooth term lo(row,col) has span 36/320.*

(0.113). When the semiparametric model was fitted to the data, fungicide treatment was significant while it was not significant in the analysis of variance. Possible reasons for that may be that the variance ratio associated with the fungicide treatment is much larger when the semiparametric model was used (see Table 5.7) and also the different F distributions used for the test statistic. The analysis of variance uses a $F_{1,2}$ and the approximate F test in the semiparametric model uses a $F_{1,329.4}$.

The fungicide treatment was measured at the main plot level, therefore it is not really appropriate to test its significance at the unit level (as we have done in the

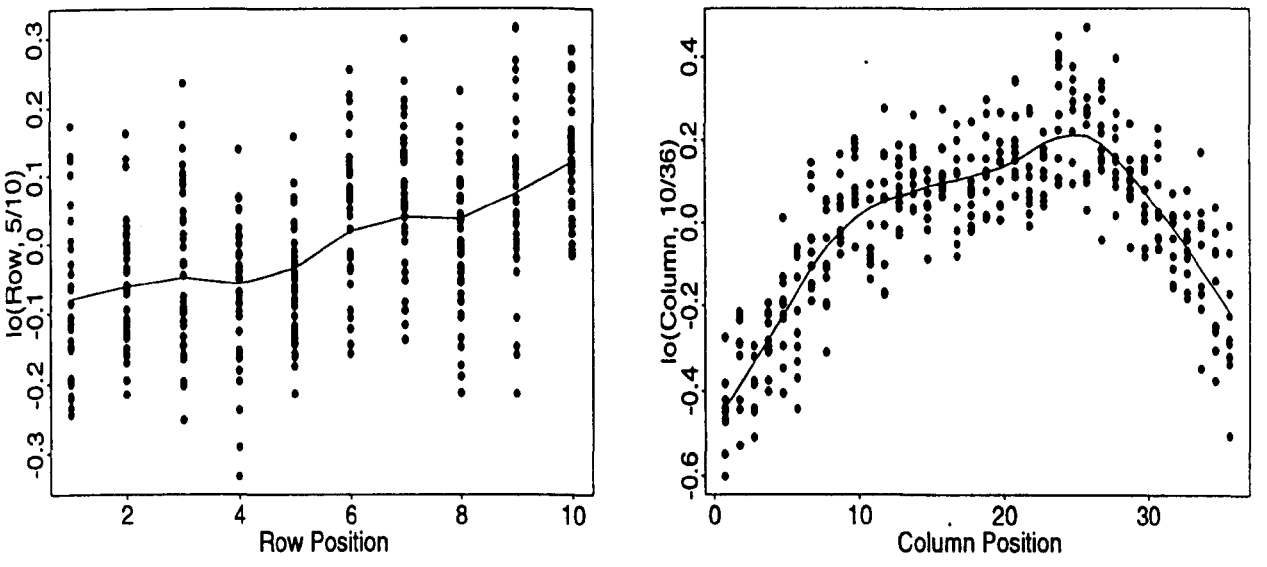

Figure 5.12: *Plot of estimated trend for row and column position.*

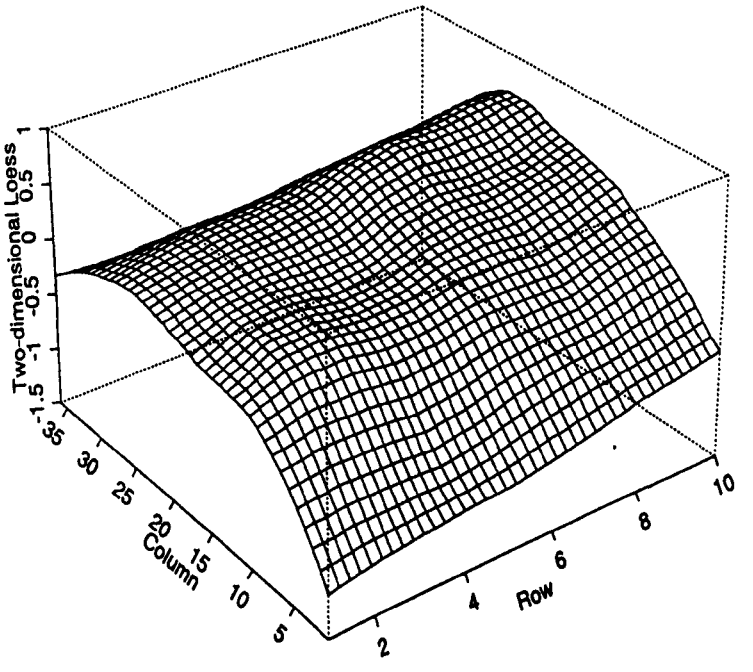

Figure 5.13: *Two-dimensional surface for unit yields of BJ trial.*

semiparametric model). Figure 5.9 showed that the trend at the sub-plot level was approximately quadratic. As an intermediate approach between analysis of variance and semiparametric modelling, we fitted a split-split-split-plot analysis of covariance (Federer and Meredith, 1992). The covariates used were a quadratic term for column position and a linear term for row position (both covariates were centred). Table 5.8 shows the analysis of covariance. The residual sum of squares for the main-plot and sub-plot stratum decreased considerably compared with the analysis of variance given in Table 5.6. This resulted in an increase of the variance ratio corresponding to the fungicide treatment from 6.9 to 96.7 (as in the semiparametric model). There is scope for further research into a semi-parametric split-plot analysis in which a smooth curve is fitted at each stratum. This would extend the split-plot analysis of covariance given in Federer and Meredith (1992) in a similar way to the nonparametric analysis of covariance (Young and Bowman, 1995).

Competition at the unit-stratum

We were particularly interested in the possibility of competition in this trial. Plots were approximately square so that competition between neighbouring plots was equally likely to occur in any direction. A model for competition between neighbouring plots (4.1) and a model for interference between neighbouring treatments (4.36) were fitted to the data, allowing for the estimation of the large-scale trend through a two-dimensional loess smoother. Three different neighbouring matrices were used, depending on the number and position of the neighbours: W_f, first order in the same row and column, W_c, first order in the same column and W_r, first order in the same row. Adjustments were made for plots in the border of the trial. We iterated between solving the equation corresponding to the modified version of the profile likelihood score (theorems 4.1 and 4.2) to estimate the competition parameter, and minimising the modified generalised cross-validation criterion (4.100) to estimate the smoothing parameter in the presence of competition. Table 5.9 summarises the results. The competition parameter was not significantly different from 0 for the model with a term for competition between neighbouring treatments. In the case of competition

Stratum	*Source of variation*	*d.f.*	*m.s.*	*v.r.*	*F Pr.*
Block	Column	1	5.481	1.14	0.479
	Residuals	1	4.816		
Main-plot	Fungicide	1	35.760	96.67	0.065
	Column	1	9.942	26.87	0.121
	Residual	1	0.369		
Sub-plot	Nitrogen	1	10.37	505.20	< 0.001
	Fung:Nit	1	0.045	2.22	0.233
	Column	1	1.492	80.53	0.003
	Residual	3	0.021		
Unit	Cultivar	29	0.895	31.18	<0.001
	Fung:Cult	29	0.075	2.64	< 0.001
	Nit:Cult	29	0.034	1.19	0.243
	Fung:Nit:Cult	29	0.032	1.13	0.298
	Row	1	1.909	66.48	< 0.001
	Column	1	0.192	6.70	0.010
	Residual	232	0.037		
Half-unit	Column	1	0.004	0.51	0.477
	Residuals	359	0.009		

Table 5.8: *Analysis of covariance of the half-unit yields of trial BJ. Column= quadratic term for column position and Row=linear term for row position.*

between neighbouring plots, there was no evidence of competition between plants in the same row. The competition parameter was significant when four neighbours (two in the same row and two in the same column) were used. The competition parameter was smaller (but also significant) for neighbouring plots in the same column. In all

Model	W_f		W_r		W_c	
	$\hat{\rho}$	$se(\hat{\rho})$	$\hat{\rho}$	$se(\hat{\rho})$	$\hat{\rho}$	$se(\hat{\rho})$
$y = X\beta + \rho Wy + f(r,c) + \epsilon$	0.15	0.053	0.040	0.040	0.110	0.043
$y = X\beta + \rho WX\beta + f(r,c) + \epsilon$	0.004	0.075	-0.005	0.047	0.006	0.054

Table 5.9: *Estimates and standard errors of the competition parameter, $\hat{\rho}$, for plot interference (top) and treatment interference (bottom) at the unit level.*

cases the competition parameter was positive, indicating an enhancement effect, i.e. plants with high yield have a positive effect in the yield of neighbouring plants. The smoothing parameter chosen by the modified generalised cross-validation criterion was 40/360 (very similar to the one chosen when the model did not include a competition effect) yielding a surface similar to the one given in Figure 5.13. Including a competition effect in the model had little effect on the average standard error of cultivar differences: it was 0.094 for a model with 4 neighbours (W_f) and 0.093 for neighbours in the same column (W_c).

Enhancement may occur in the context of plant diseases, however it is difficult to justify when the response measured is yield. It is possible that competition and local correlation are being confused.

Analysis of half-unit yields

Here, we fit a semiparametric model to the yield at the half unit stratum. For such a spatial analysis it is debatable whether we should be basing the analysis on the unit yields or half-unit yields. It is instructive to compare the two analyses.

The span selected for the two-dimensional loess was 10/720, while the best model with two one-dimensional smoothers included a linear term for row position and a loess smoother with span 5/72 for column position. Figure 5.14 shows a plot of the fitted trend for the additive and two-dimensional smoother. The fitted trend is not smooth in either of the models, indicating that we might be over-fitting the

model (even when the smoothing parameter was chosen by the modified AIC). Both

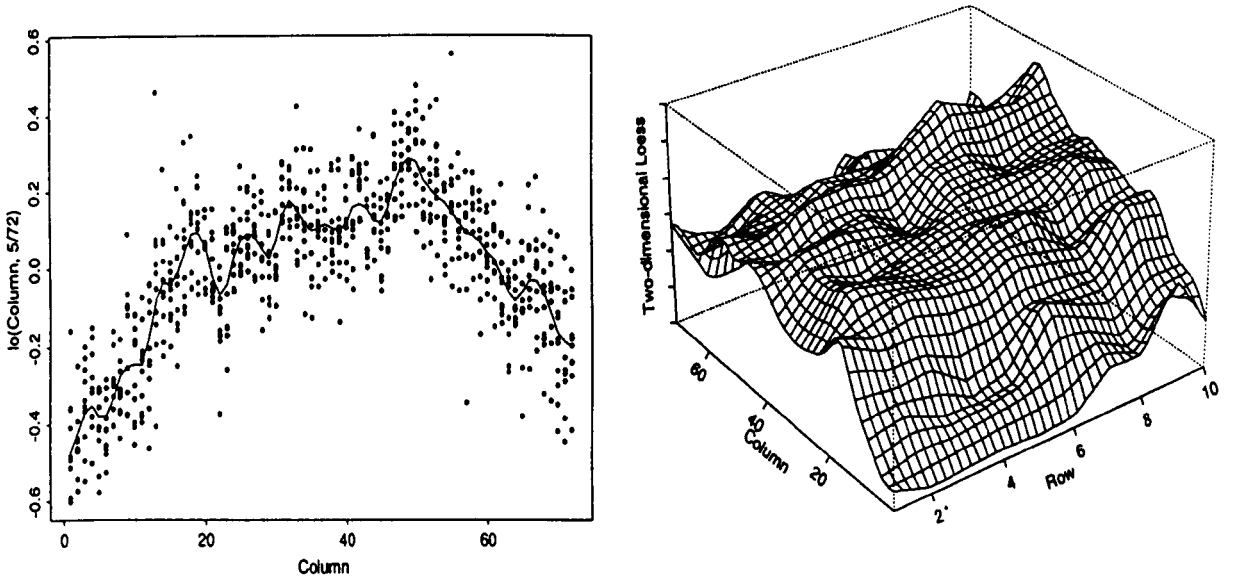

Figure 5.14: *Plot of the fitted loess smoother for column position and two-dimensional smooth surface.*

the approximate and bootstrapped F test agreed in selecting the model with a two-dimensional smoother as more appropriate. Compared with the residual mean square at the half-unit level, all main effects, the interaction between cultivar and fungicide and the three-way interaction were significant (the average standard error of cultivar differences was 0.035). However, we need to be cautious in drawing any conclusions about significance of treatments effects since no treatment was measured at the half-unit level.

One of the assumptions underlying the theory of additive models is that the errors are uncorrelated. When this condition is not satisfied, the use of the criteria of section 5.2.4 for the selection of the smoothing parameter leads to an under-smoothed curve (Altman, 1990; Diggle and Hutchinson, 1989; Wang, 1998b). An important issue is to identify the biological process that is inducing the correlation, and to select the appropriate model to estimate it. There are two possible ways to interpret the presence of correlation in the residuals: due to local spatial variation (as assumed in models

such as those given in Wang, 1998b; Verbyla et al., 1998) or due to interference between plots which is the approach we take here. A model for plot competition, (4.1), was fitted to the data (see Table 5.10). (A model for competition between neighbouring treatments, (4.36), does not make sense since each variety would have itself as one of its neighbours). The competition parameter for model (4.1) was significant

Model	W_f		W_r		W_c	
	$\hat{\rho}$	$se(\hat{\rho})$	$\hat{\rho}$	$se(\hat{\rho})$	$\hat{\rho}$	$se(\hat{\rho})$
$y = X\beta + \rho Wy + f(r,c) + \epsilon$	0.23	0.04	0.14	0.03	0.07	0.032

Table 5.10: *Estimates and standard errors for the competition parameter for a model with competition between neighbouring half-units.*

for all three neighbouring matrices. It is worth noting that for half-unit yields the competition parameter for plots in the same row was significant (it was not when unit yields was used as response), this is almost certainly due to the fact that one of the neighbours is the same cultivar. This induces a positive correlation which is being confused with competition. We have analysed in detail the model with neighbouring matrix W_f and competition parameter 0.23. Fitting a term for competition has a major impact on the selection of the span. The smoothing parameter for the two-dimensional loess increased from 10/720 to 46/720, which gave a smoother surface (see Figure 5.15), similar to the one obtained when the semiparametric model was based on the unit yields (Figure 5.13). The average standard error of cultivar difference was 0.0371, larger than the one obtained with the semiparametric model (0.0354) with no competition term and span 10/720.

We would normally assume that the extra spatial information provided by the half-unit yields in the BJ trial would be helpful. However, it yielded other problems: there was local correlation which resulted in overfitting and an uneven surface. When the competition model was fitted, plots had the same cultivar as a neighbour, and this resulted in a significant positive value of the competition parameter. However, evidence for competition or local correlation between plots in the same row disappeared when

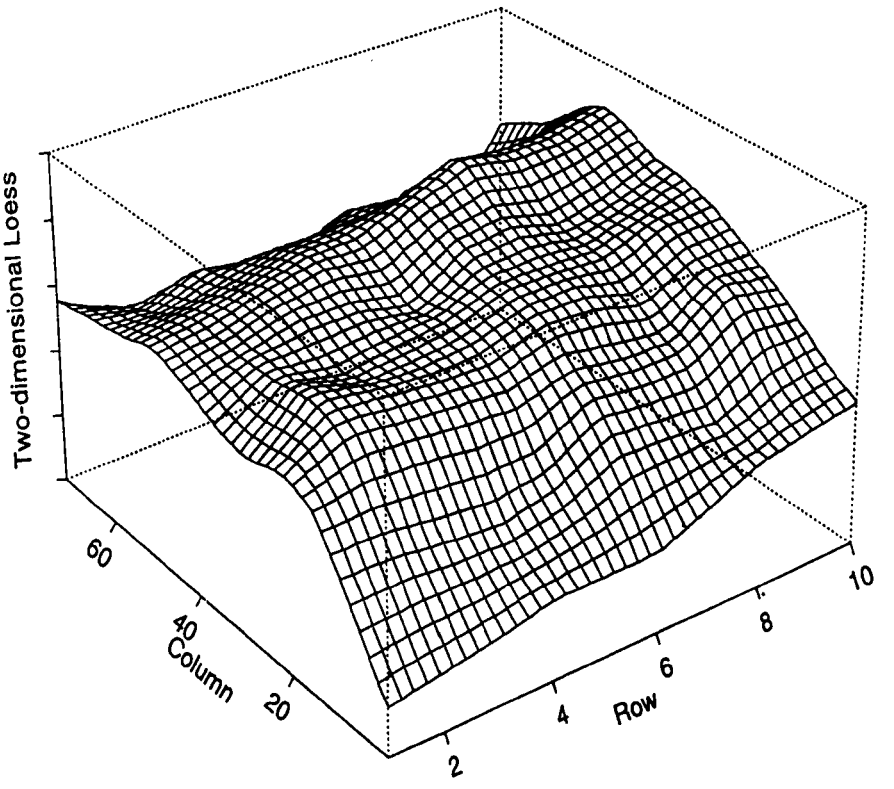

Figure 5.15: *Two-dimensional loess for yields of BJ trials after adjusting for competition.*

unit yields were used for the analysis. Hence, using half-unit yields led to confusion in identifying the source of correlation rather than improving the fit of the model.

Remarks for trial BJ

Remark 1 It is not possible to integrate the modified profile log-likelihood score; therefore we do not have an expression for the modified profile likelihood, and models with and without competition effect cannot be compared. Intuitively, Figure 5.13 at the unit level gives the best representation of the trend and is a more plausible trend in yield than the one shown in Figure 5.14 at the half-unit level. There is no apparent reason why the yield in a plot would enhance the yield in the neighbouring plot in this

trial, so we are inclined to conclude that the effect modelled here as competition is actually local correlation. Although the interference model might not be appropriate it helped to recover the smooth trend at the half-unit level (see Figure 5.15). An alternative would be to use models of the type proposed by Verbyla et al. (1998) and use an $AR \times AR$ process to estimate the local trend. However, at the present moment these models do not include two-dimensional smoothers. It is our opinion, from the analysis of several trials, that in many cases a non-additive trend is present in field trials. Therefore, we consider models given in Verbyla et al. (1998) incomplete. This leaves open the problem of estimating local correlation when a two-dimensional smoother is used to account for the large-scale trend.

Remark 2 The significance of the treatments did not change after including a term for competition. In the presence of a competition effect, a two-dimensional surface also gave a better fit to the data than two one-dimensional terms.

Remark 3 The approximate calculation of the standard errors given in section 2.4 proved to be very efficient for trial BJ. The number of operations necessary to calculate SX was much smaller than the number necessary to calculate S (a 720×720 matrix). The smoothing parameter was quite small (10/720 or 42/720), therefore we would expect S to be close to a symmetric matrix. Hence, we can use the approximation confidently.

Chapter 6

Conclusions and further work

6.1 Conclusions

The aim of this project was to develop a joint approach to the estimation of spatial trends and competition effects in agricultural field trials. First, we studied in detail the use of additive models as a data-driven modelling method for estimating spatial trends. Second, we reviewed the methodology already developed for competition models, and finally we approached the task of combining smoothing and competition in a single model. Each of these areas revealed some unsolved questions which have been the focus of our research, as we summarise here.

6.1.1 Additive and semiparametric models

Additive models were introduced by Hastie and Tibshirani (1986) and most of the theoretical results are given in Buja et al. (1989). However most of these results were constrained to the case of two smooth terms and results on semiparametric models included only one smooth term. We interpreted semiparametric models as a particular case of an additive model. Therefore, by giving explicit expressions for the estimated smooth terms in an additive model with any number of smooth terms, we have obtained a closed form for the estimates of the linear part of the model and their standard errors (which could not be obtained by using the back-fitting algorithm). A more straightforward approach than the one given in Buja et al. (1989) to the

conditions for existence and uniqueness of the solutions of the normal equations was also obtained.

In the case of field trials, the estimation of the treatments and their standard errors is of most interest. The approximation we have proposed for the calculation of the standard errors allowed us to handle large data sets for which exact results would be computationally very demanding. The properties derived for the form of the hat matrix of an additive model with any number of smoothers revealed that for smoothing splines (or any other symmetric smoother) the approximation is exact and with loess it gives good estimates of the standard errors. (See also Durban et al., 1999.)

Most of the recent work in spatial modelling for agricultural trials is restricted to experiments with a single treatment factor. Semiparametric additive models (SAMs) can be used in both single and multi-factor experiments, although further research is needed for designs with several strata. SAMs provide a flexible framework for identifying underlying trends in field trials. They generally improve the precision of the treatment estimates and they enable spatial trends to be easily visualised, in contrast to models such as the ARIMA models of Cullis and Gleeson (1991). The graphical methods proposed for the selection of the spatial component give an intuitive image of the underlying trend; problems with model selection and tests of treatment significance may be overcome by using the bootstrapped F test.

6.1.2 Profile likelihood

Profile likelihood was the method previously used for the estimation of the competition parameter in a model with *treatment+block+competition.* This method gave biased estimates of the variance parameters and a small simulation study showed that, in some cases, the competition parameter was also biased. In the competition model of Besag and Kempton (1986), the competition parameter is present in the mean and variance. We do not see how a REML style argument can be used in this example, since the purpose of REML is only to adjust the estimates of the variance

parameters. We have taken the approach of McCullagh and Tibshirani (1990); in their paper, a first order approximation to the adjustments of the profile likelihood score was derived. We have shown that their adjustments are exact for the class of normal regression models: $y \sim \mathcal{N}(X(\psi)\lambda, \Sigma(\psi))$ (which allows parameters of interest, ψ, in both the mean and the variance). The matrix notation that we used simplifies the tensor approach taken by McCullagh and Tibshirani (1990). An important result obtained was that the adjusted profile likelihood was equivalent to REML when the parameters of interest were present only in the variance; when the parameters are present in both mean and variance, the estimates of the variance parameters are also REML style estimates. Although the motivation for the development of this method was the competition model, the class of regression models covers a much wider range of models (see Durban and Currie, 1998).

6.1.3 Semiparametric models and competition

The non-parametric nature of the smoothing approach did not fit into the profile likelihood context. We overcame this difficulty by *"parameterising"* the non-parametric component of a semiparametric model (in the way suggested in Speckman, 1988). There are several reasons to use this method with confidence: the estimates of the treatment and smooth terms are the same as the ones in an ordinary SAM but with the response adjusted for the competition effect, and it yields the intuitive extension of the results given in Besag and Kempton (1986). As in the case of semiparametric models with no competition effect, inference is done on the basis of a known smoothing parameter. Competition induces correlation in the errors, and thus data are under-smoothed if the usual criteria for the selection of the smoothing parameter are used. We have adapted the GCV criterion proposed by Altman (1990) to choose the smoothing parameter in the presence of competition. However, this method also recovered the correct trend when the source of correlation was not competition, but local correlation among the residuals (BJ trial in chapter 5). The competition and smoothing parameter are estimated by iterating between the adjusted profile likelihood and the modified GCV.

Recent developments in the field of smoothing have proved that cubic smoothing splines admit a mixed model decomposition. This allows us to account for spatial trend (in one direction) and competition effects in a fully parametric context. The smoothing and competition parameter can be estimated simultaneously. This resulted, however, in high correlation between these two parameters, and suggests that, in some cases, there might be problems with the identifiability of both effects.

6.2 Further work

6.2.1 Semiparametric split-plot analysis

Spatial models based on the *trend+error* decomposition (see chapter 1) have been used in randomised block designs with a single treatment factor. When a design with multiple strata is used in an experiment, none of the spatial models developed until now takes into account the design. In a split-plot design, treatments (main-plot, sub-plot, etc) are randomised at different strata, therefore the intuitive approach would be to test for the significance of each treatment at the stratum in which it has been measured. Semiparametric models can be confidently used in multi-factor experiments with a single stratum or in multi-stratum experiments to test for the significance of the treatments at the bottom stratum (as for the cultivar treatment in the examples in chapter 5). It is debatable whether it is justified to ignore the split-plot design when fitting a smooth trend and test all treatments as if there was only one stratum. A possible alternative would be the extension of the split-plot analysis of covariance to a multi-stratum semiparametric model in which the linear function for the covariate at each stratum is substituted by a smooth curve or surface. Given a covariate Z, the model for split-plot analysis of covariance is written as

$$y_{ijk} = \mu + \rho_i + \tau_j + \theta_1(\bar{Z}_{ij.} - \bar{Z}_{...}) + \delta_{ij} + \alpha_k + (\tau\alpha)_{jk} + \theta_2(Z_{ijk} - \bar{Z}_{ij.}) + \epsilon_{ijk},$$

where μ is the overall mean, ρ is the randomised block effect, τ is the main-plot treatment, α is the sub-plot treatment, δ is a random main-plot error and ϵ is the random split-plot error. The idea behind the non-parametric split-plot analysis would be to

replace the linear functions of the covariate Z by smooth functions of plot position at each stratum. There are, of course, implications for the design: to be able to fit a curve we would need more replication at each stratum. Careful planning of the position of main-plots and sub-plots will also be necessary to allow for the estimation of a two-dimensional trend.

Another possible approach to the semiparametric split-plot analysis is to use the mixed model decomposition of cubic splines and use similar ideas to the ones in the ANOVA with mixed-effects models. Some related work (Brumback and Rice, 1998) in this context has recently appeared.

6.2.2 Profile likelihood

The use of the profile likelihood and the closed form derived for the adjustments have opened further lines of research. When the parameters of interest are present in the mean and variance, the profile likelihood score cannot be integrated, and thus a closed form for the adjusted profile likelihood cannot be obtained. We proposed in chapter 3 an approximate method to obtain it. However, methods to compare models when the adjusted likelihood is not available need further research.

The bias and variance adjustments depend on the trace of matrices which may be quite large in some cases. Further investigation on approximations for the bias adjustment is needed. Computation of the adjustment information matrix might be simplified by extending the average information REML of Gilmour et al. (1995) to the class of regression models we presented in chapter 3.

Kenward and Roger (1997) gave improved estimates for the standard errors of fixed effects estimated via REML. It would be of interest to compare them with standard errors obtained from the adjusted profile likelihood when the regression parameters are taken as parameters of interest. This method could be also applied to the standard errors of fixed effects for models in which the mixed-model decomposition of

splines is adopted.

6.2.3 Smoothing and mixed models

Several authors (Speckman, 1988; Severini and Wong, 1992) have studied the asymptotic properties of the method based on partial residuals (proposed by Speckman, 1988) and concluded that in some cases the estimates of the regression part of the model are preferable to the estimates given in Green et al. (1985). Further investigation is needed to show whether the same argument follows in the extension of Speckman's method to semiparametric models with more than one smooth terms that we proposed in chapter 2.

One of the main conclusions drawn from our work is that a spatial model for the analysis of agricultural field trails should be able to account for two-dimensional trends and local correlation, and correlation and smoothing parameters should be estimated simultaneously. This would be achieved if a mixed-model approach is used; however further work is still needed in this field. Models such as those given in Verbyla et al. (1998) account for local correlation and one-dimensional trends. These models may be improved if a mixed model decomposition for the thin-plate spline is found and incorporated in the model. The smoothing approach that we use here may also be improved if an approximate mixed-model decomposition for loess is found. These models could then easily be extended to estimate competition effects.

Bibliography

Akaike, H. (1973). Information theory and an extention of the maximum likelihood principle. In Petrof, B. and Csàki, editors, *Second International Symposium on Information*, pages 267–281, Akademia Kiadó, Budapest.

Altman, N. (1990). Kernel smoothing of data with correlated errors. *J. Am. Statist. Assoc.*, 85:749–759.

Azaïs, J.-M., Bailey, R., and Monod, H. (1993). A catalogue of efficient neighbour-designs with border plots. *Biometrics*, 49:1252–1261.

Bailey, R. (1984). Quasi-complete latin squares: Construction and randomisation. *J. R. Statist. Soc. B*, 46:323–334.

Barndorff-Nielsen, O. (1986). Inference on full or partial parameters, based on the standarized log-likelihood ratio. *Biometrika*, 73:307–322.

Bartlett, M. (1938). The approximate recovery of information from field experiments with large blocks. *Journal of Agricultural Sciences*, 28:418–427.

Bartlett, M. (1978). Nearest neighbour models in the analysis of field experiments. *J. R. Statist. Soc. B*, 40:147–174.

Besag, J. (1984). A method for the analysis of field experiments based on first differences. In *Spatial Methods in Field Experiments. Proceedings of the Biometric Society Workshop*, pages 9–11, University of Durham.

Besag, J. and Kempton, R. (1986). Statistical analysis of field experiments using neigbouring plots. *Biometrics*, 42:231–251.

Binns, M. and Jui, P. (1985). A simulation study of the nearest neighbour analysis method of Papadakis. *Communications in Statistics - Simulation and Computation*, 14:159–172.

Bowman, A. and Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations.* Oxford Statistical Science Series. Oxford University Press, Oxford.

Brownie, C. and Gumpertz, M. L. (1997). Validity of spatial analysis for large field trials. *Journal of Agric., Bio. and Env. Stats.*, 2:1–23.

Brumback, B. and Rice, J. (1998). Smoothing spline models for the analysis of nested and crossed samples of curves. *J. Am. Statist. Assoc.*, 93:961–994.

Buckley, M., Eagleson, G., and Silverman, B. (1988). The estimation of residual variance in nonparametric regression. *Biometrika*, 75:183–199.

Buja, A., Hastie, T., and Tibshirani, R. (1989). Linear smoothers and additive models (with discussion). *Annals of Statistics*, 17:453–555.

Carver, T. and Griffiths, E. (1981). Relationship between powdery mildew infection, green leaf and grain yield of barley. *Ann. Appl. Biol.*, 99:255–266.

Chu, C.-K. and Marron, J. (1991). Choosing a kernel regression estimator (with discussion). *Statistical Science*, 6:404–436.

Cleveland, W. (1979). Robust locally-weighted regression and smoothing scatterplots. *J. Am. Statist. Assoc.*, 74:829–836.

Cleveland, W. (1994). Coplots, non-parametric regression, and conditionally parametric fits. In Anderson, T., Fang, K., and Olkin, I., editors, *IMS Lecture Notes*, volume 24, pages 21–36. Hayward,California.

Cleveland, W., Devlin, S., and Grosse, E. (1988). Regression by local fitting. *Journal of Econometrics*, 37:87—114.

Cleveland, W. and Loader, C. (1996). Smoothing by local regression: Principles and methods. In Härdle, W. and Schimek, M., editors, *Statistical Theory and Computational Aspects of Smoothing*, pages 10–49. Physica-Verlag, New York.

Cleveland, W. S. (1993). *Visualizing Data.* Hobart Press, Summit, New Jersey.

Cleveland, W. S. and Devlin, S. J. (1988). Locally weighted regression: An approach to regression analysis by local fitting. *J. Am. Statist. Assoc.*, 83:596–610.

Connolly, T. (1992). *Competition in Variety Trials.* PhD thesis, Heriot-Watt University.

Connolly, T., Currie, I., Bradshaw, J., and McNicol, J. (1993). Inter-plot competition in yield trials of potatoes (Solarum tuberosum L.) with single-drill plots. *Ann. Appl. Biol.*, 123:367–377.

Cooper, D. and Thompson, R. (1977). A note on the estimation of the parameters of the autoregressive-moving average process. *Biometrika*, 64:625–628.

Cox, D. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference. *J. R. Statist. Soc. B*, 49:1–39.

Cox, D. and Reid, N. (1993). A note on the calculation of adjusted profile likelihood. *J. R. Statist. Soc. B*, 55:467–471.

Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of cross-validation. *Numer. Math.*, 31:377–403.

Cullis, B. R. and Gleeson, A. C. (1991). Spatial analysis of field experiments. An extension to two dimensions. *Biometrics*, 47:1449–1460.

Cuzick, J. (1992). Semiparametric additive regression. *J. R. Statist. Soc. B*, 54:831–843.

Devlin, S. (1986). Locally-weighted multiple regression: Statistical properties and its use to test for linearity. Research technical memorandum, Bell Communications.

Diggle, P. and Hutchinson, M. (1989). On spline smoothing with autocorrelated errors. *Austral. J. Statist.*, 31:166–182.

Draper, N. and Guttman, I. (1980). Incorporating overlap effects from neighbouring units into response surface models. *Applied Statistics*, 29:128–134.

Durban, M. and Currie, I. (1998). Adjustment of the profile likelihood for a class of normal regression models. *Submitted.*

Durban, M., Hackett, C., and Currie, I. (1999). Approximate standard errors in semi-parametric additive models. *To appear in Biometrics.*

Eccleston, J. and Chan, B. (1998). Design algorithms for correlated data. In Payne, R. and Green, P., editors, *Proceedings in Computational Statistics*, pages 41–52, Bristol. Physica-Verlag.

Eubank, R. (1984). The hat matrix for smoothing splines. *Statistics and Probability Letters*, 2:9–14.

Federer, W. and Basford, K. (1991). Competing effects designs and models for two-dimensional field arrangements. *Biometrics*, 47:1461–1472.

Federer, W. and Meredith, M. (1992). Covariance analysis for split-plot and split-block designs. *The American Statistician*, 46:155–162.

Freeman, G. (1979a). Complete latin squares and related experimental designs. *J. R. Statist. Soc. B*, 41:253–262.

Freeman, G. (1979b). Some two-dimensional designs balanced for nearest neighbours. *J. R. Statist. Soc. B*, 41:88–95.

Gampe, J. (1998). Trend or correlation? - Attributing smoothness in nonparametric regression. In Marx, B. and Friedl, H., editors, *Statistical Modeling. Proceedings of the 13th International Workshop on Statistical Modeling*, pages 216–221, New Orleans, Louisiana.

Gasser, T. and Müller, H.-G. (1979). Kernel estimation of regression functions. In Gasser, T. and Rosenblatt, M., editors, *Smoothing Techniques for Curve Estimation*, pages 23–68, Heidelberg. Springer-Verlag.

Gasser, T. and Müller, H.-G. (1984). Estimating regression functions and their derivatives by the kernel method. *Scandinavian Journal of Statistics*, 11:171–185.

Genstat-5, C. (1995). Genstat 5 Release 3.2. Reference Manual Supplement, IARC-Rothamsted, Harpenden, Hertfordshire.

Gilmour, A., Cullis, B., and Verbyla, A. (1997). Accounting for natural and extraneous variation in the analysis is of field experiments. *J. Agric., Bio. and Env. Stat.*, 2:269–293.

Gilmour, A. R., Thompson, R., and Cullis, B. R. (1995). Average information REML: an efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics*, 51:1440–1450.

Gleeson, A. and Cullis, B. (1987). Residual maximum likelihood (REML) estimation of a neighbour model for field experiments. *Biometrics*, 43:277–288.

Gogel, B., Cullis, B., and Verbyla, A. (1995). REML estimation of multiplicative effects in multi-environment variety trials. *Biometrics*, 51:745–749.

Green, P. (1985). Linear models for field trials, smoothing and cross-validation. *Biometrika*, 72:527–537.

Green, P. (1987). Penalized likelihood for general semi-parametric regression models. *International Statistical Review*, 55:245–259.

Green, P., Jennison, C., and Seheult, A. (1985). Analysis of field experiments by least squares smoothing. *J. R. Statist. Soc. B*, 47:299–315.

Green, P. and Silverman, B. (1994). *Nonparametric Regression and Generalized Linear Models.* Monographs on Statistics and Applied Probability. Chapman & Hall, London.

Hackett, C. and Newton, A. (1995). Modelling spatial trend in barley field trials using generalized additive models. *Aspects of Applied Biology*, 43:59–66.

Hart, J. (1991). Kernel regression estimation with time series errors. *J. R. Statist. Soc. B*, 53:173–187.

Harville, J. (1974). Bayesian inference for variance components using error contrasts. *Biometrika*, 61:383–385.

Hastie, T. and Loader, C. (1993). Local regression: Automatic kernel carpentry (with discussion). *Statistical Science*, 8:120–143.

Hastie, T. and Tibshirani, R. (1986). Generalized additive models. *Statistical Science*, 1:297–318.

Hastie, T. and Tibshirani, R. (1987). Generalized additive models: Some applications. *J. Am. Statist. Assoc.*, 82:371–386.

Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models.* Monographs on Statistics and Applied Probability. Chapman and Hall, London.

Hastie, T. and Tibshirani, R. (1998). Bayesian backfitting. Technical report, Department of Statistics and Division of Biostatistics, Stanford University, Stanford, California.

Heckman, N. (1986). Spline smoothing in a partly linear model. *J. R. Statist. Soc. B*, 48:244–248.

Henderson, C., Kempthorne, O., Searle, S., and Von-Krosigk, C. (1959). Estimation of environmental and genetic trends from records subject to culling. *Biometrics*, 15:192–218.

Hurvich, C., Simonoff, J., and Tsai, C.-L. (1998). Smoothing parameter selection in nonparametric regression using an improved Akaike information criteria. *J. R. Statist. Soc. B*, 60:271–293.

Hurvich, C. and Tsai, C.-L. (1989). Regression and time-series model selection in small samples. *Biometrika*, 76:297–307.

Jenkyn, J., Bainbridge, A., Dyke, G., and Todd, A. (1979). An investigation into inter-plot interactions, in experiments with mildew on barley, using balanced designs. *Ann. Appl. Biol.*, 92:11–28.

Kalbfleisch, J. and Sprott, D. (1970). Application of likelihood methods to models involving large number of parameters. *J. R. Statist.Soc. B*, 32:175–194.

Kauermann, G. (1997). A profile likelihood adjustment in semiparametric models. Technical report, Institute of Quantitative Models, Technical University, Berlin.

Kempton, R. (1982). Adjustment for competition between varieties in plant breeding trials. *Journal of Agricultural Sciences, Cambridge*, 98:599–611.

Kempton, R. (1984). Nearest neighbour analysis and inter-plot competition. In *Spatial Methods in Field Experiments. Proceedings of the Biometric Society Workshop*, pages 34–43, University of Durham.

Kempton, R. (1985). Statistical models for interplot competition. *Aspects of Applied Biology*, 10:111–120.

Kempton, R. and Lockwood, G. (1984). Inter-plot competition in variety trials of field beans. *Journal of Agricultural Sciences, Cambridge*, 103:293–302.

Kempton, R., Seraphin, J., and Sword, A. (1994). Statistical analysis of two-dimensional variation in variety yield trials. *Journal of Agricultural Sciences, Cambridge*, 122:335–342.

Kenward, M. and Roger, J. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, 53:983–997.

Khatri, G. (1966). A note on a manova model applied to problems in growth curves. *Ann. Inst. Stat. Math.*, 18:75–78.

Lill, W., Gleeson, A., and Cullis, B. (1988). Relative accuracy of a neighbour method for field trials. *Journal of Agricultural Sciences, Cambridge*, 111:339–346.

Loader, C. (1995). Old faithful erupts: Bandwidth selection reviewed. Technical report, AT & T Bell Laboratories.

Macaskill, G. T. (1993). A note on adjusted profile likelihoods in non-linear regression. *J. R. Statist. Soc. B*, 55:125–131.

Martin, R. (1990). The use of time-series models and methods in the analysis of agricultural field trials. *Communications in Statistics-Theory and Methods*, 19:55–81.

McCullagh, P. (1987). *Tensor Methods in Statistics.* Monographs on Statistics and Applied Probability. Chapman and Hall, London.

McCullagh, P. and Tibshirani, R. (1990). A simple method for the adjustment of profile likelihoods. *J. R. Statist. Soc. B*, 52:325–344.

Müller, H.-G. (1987). Weighted local regression and kernel methods for non-parametric regression. *J. Am. Statist. Assoc.*, 82:231–238.

Nadayara, E. (1964). On estimating regression. *Theory of Probability and Its Applications*, 10:186–190.

Nelder, J. (1969). The combination of information in generally balanced designs. *J. R. Statist. Soc. B*, 30:303–311.

Opsomer, J.-D.and Ruppert, D. (1997). Fitting a bivariate additive model by local polynomial regression. *Annals of Statistics*, 25:186–211.

Papadakis, J. (1937). Methodé statistique pour des expériences sur champ. *Bulletin de l'Institut d'Amélioration des Plantes à Salonique*, 23.

Patterson, H. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, 58:545–554.

Patterson, H. and Williams, E. (1976). A new class of resolvable incomplete block designs. *Biometrika*, 63:83–92.

Pearce, S. (1957). Experimenting with organisms as blocks. *Biometrika*, 41:434–449.

Pearce, S. (1983). *The Agricultural Field Experiment.* John Wiley & Sons, New York.

Rice, J. (1986). Convergence rates for partially splined models. *Statistics and Probability Letters*, 4:203–208.

Ruppert, D., Sheather, S., and Wand, M. (1995). An effective bandwidth selector for local least squares regression. *J. Am. Statist. Assoc.*, 90:1257–1270.

Russel, K. and Eccleston, J. (1987a). The construction of optimal balanced incomplete block designs when adjacent observations are correlated. *Austral. J. Statist.*, 29:84–90.

Russel, K. and Eccleston, J. (1987b). The construction of optimal incomplete block designs when observations within a block are correlated. *Austral. J. Statist.*, 29:293–302.

Schimeck, M. G. (1997). Non- and semiparametric alternatives to generalized linear models. *Computational Statistics*, 12:173–191.

Schott, J. R. (1997). *Matrix Analysis for Statistics.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Son, New York.

Severini, T. and Wong, W. (1992). Profile likelihood and conditionally parametric models. *Annals of Statistics*, 20:1768–1802.

Silverman, B. (1984). Spline smoothing: the equivalent variable kernel method. *Annals of Statistics*, 12:898–916.

Silverman, B. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting (with discussion). *J. R. Statist. Soc. B*, 47:1–52.

Simonoff, J. (1996). *Smoothing Methods in Statistics.* Springer Series in Statistics. Springer-Verlag, New York.

Simonoff, J. and Tsai, C.-L. (1999). Semiparametric and additive model selection using an improved AIC criterion. *Journal of Computational and Graphical Statistics*, To Appear.

Speckman, P. (1988). Kernel smoothing in partial linear models. *J. R. Statist. Soc. B*, 50:413–436.

Stafford, J. (1996). A robust adjustment of the profile likelihood. *Annals of Statistics*, 24:336–352.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions (with discussion). *J. R. Statist. Soc. B*, 36:111–147.

Talbot, M., Milner, A., Nutkins, M., and Law, J. (1995). Effect of interference between plots on yield performance in crop variety trials. *Journal of Agricultural Sciences, Cambridge*, 124:335–342.

Tunnicliffe Wilson, G. (1989). On the use of marginal likelihood in time series model estimation. *J. R. Statist. Soc. B*, 51:15–27.

Verbyla, A., Cullis, B., Kenward, M., and Welham, S. (1998). The analysis of designed experiments and longitudinal data using smoothing splines. *J. R. Statist. Soc.*, To Appear.

Wahba, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. *J. R. Statist. Soc. B*, 40:364–372.

Wahba, G. (1983). Bayesian "confidence intervals" for the cross-validated smoothing spline. *J. R. Statist. Soc. B*, 45:133–150.

Wang, Y. (1998a). Mixed effects smoothing spline analysis of variance. *J. R. Statist. Soc. B*, 60:159–174.

Wang, Y. (1998b). Smoothing spline models with correlated errors. *J. Am. Statist. Assoc.*, 93:341–348.

Wiancko, A. (1914). Use and management of check plots in soil fertility investigations. *J. Am. Soc. Agro.*, 6:122–124.

Wilkinson, G., Eckert, S., Hancock, T., and Mayo, O. (1983). Nearest Neighbour (NN) analysis of field experiments (with discussion). *J. R. Statist. Soc. B*, 45:151–211.

Williams, E. (1986). Row and column designs with continuous replicates. *Austral. J. Statist.*, 28:154–163.

Williams, E. and Jones, J. (1996). Row-column factorial designs for use in agricultural field trials. *Applied Statistics*, 45:39–46.

Williams, R. (1952). Experimental designs for serially correlated observations. *Biometrika*, 39:151–167.

Yates, F. (1936). A new method of arranging variety trials involving a large number of varieties. *Journal of Agricultural Sciences, Cambridge*, 26:424–455.

Young, S. and Bowman, A. (1995). Non-parametric analysis of covariance. *Biometrics*, 51:920–931.

Zimmerman, D. and Harville, D. (1989). On the unbiasedness of the Papadakis estimator and other nonlinear estimators of treatment contrasts in field-plot experiments. *Biometrika*, 76:153–159.

Zimmerman, D. and Harville, D. (1991). A random field approach to the analysis of field-plot experiments and other spatial experiments. *Biometrics*, 47:223–239.